

日本の研究者によるデータ公開の実践状況と認識

著者	池内 有為
発行年	2019
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2019
報告番号	12102乙第2937号
URL	http://doi.org/10.15068/00158168

日本の研究者によるデータ公開の実践状況と認識

2019年9月

池 内 有 為

日本の研究者によるデータ公開の実践状況と認識

筑波大学

図書館情報メディア研究科

2019年9月

池 内 有 為

概要

日本の研究者によるデータ公開の実践状況と認識

【研究背景】

地球科学や天文学，ゲノミクス，素粒子物理学などの分野では，観測データや採取コストの高いデータを共有することによって研究の発展を遂げてきた。2000 年ごろからはインターネットを介して世界規模でデータの共有や共同研究を行う e-Science が多くの分野に広がり，各国の政府や国際組織によって研究データの公開に関する政策が打ち出されるようになった。その背景には，先行分野のような研究の効率化や研究開発費の節減に加えて，研究不正への対策，研究成果の市民への還元といった社会的側面，および情報通信技術の発達と普及やデータ集約型科学の隆盛といった技術的側面がある。間接的には，産業界におけるビッグデータの収集とそれを分析するデータサイエンティストの活躍や，政府機関によるオープンデータの広がりやオープンデータを活用した新たなサービスの開発などによって，社会全体でデータの利活用が盛んになっていることも影響していると考えられる。さらに，分野や地域を超えて研究データを利活用することによって，異分野データの融合による新たな発見，産学官と市民の連携による社会課題の解決，イノベーションの創出による経済的効果，市民参加型科学（シチズンサイエンス）の拡大，教育での活用など，さまざまな成果を生み出すことも期待されている。

こうした動向をふまえて，公的資金による研究の成果である論文とデータを公開して市民が自由にアクセスできるようにするためのオープンサイエンス政策が，経済協力開発機構（OECD）や G7 科学大臣会合などの主導によって世界規模で広がりつつある。日本においても 2015 年ごろから議論が盛んになり，2016 年に公開された「第 5 期科学技術基本計画」ではオープンサイエンスの推進が学術政策の課題の 1 つとして位置づけられている。

一方，データ公開を実現するためには，インフラの整備や研究者の懸念など多くの課題が存在する。また，オープンサイエンス政策は国や分野による特性をふまえて適切に進める必要があると指摘されている。日本の研究者は研究開発費や研究時間の減少によって，国際競争力が低下していると指摘されていることから，オープンサイエンス政策の推進にあたっては，研究活動を阻害しないよう慎重な検討が必要であると考えられる。しかし現状では，データ公開に関する課題や分野による差異，および日本の研究者によるデータ公開の状況は十分に明らかにされていない。

【目的】

そこで本研究は，日本の研究者によるデータ公開の実践状況およびデータ公開に対する認識を明らかにすることを目的とする。研究者の認識は，データ公開の動機や障壁，公開データの利用経験といった内的要因から明らかにすることとした。また，データ公開を推進する

とされている学術雑誌のポリシーを外的要因として調査し、日本の研究者によるデータ公開の状況との比較を行うこととした。研究課題 (Research Questions, RQ) として, [RQ1]日本の研究者はどの程度データを公開しているのか, [RQ2]日本の研究者によるデータ公開の内的要因は何か, [RQ3]データ公開の外的要因はどのような状況にあるのか, この外的要因と日本の研究者によるデータ公開状況に乖離はあるのか, という 3 点を設定した。

【方法】

[RQ1]と[RQ2], すなわち日本の研究者によるデータ公開の状況と内的要因を明らかにするために, 2016 年 11 月から 12 月にかけて質問紙調査を実施した。また, [RQ3]データ公開の外的要因の状況を明らかにするために, 2014 年 4 月から 6 月にかけて学術雑誌のデータ公開ポリシーを調査して, 分野別にデータ公開の要求状況をマクロに捉えることとした。その上で, 質問紙調査の結果 (日本の研究者によるデータ公開率) と, 学術雑誌のポリシー調査の結果 (データ公開の要求状況) を分野別に比較した。

質問紙調査は, 文部科学省科学技術・学術政策研究所 (NISTEP) が運営する科学技術専門家ネットワークに登録されている研究者 1,983 名を対象として実施した。科学技術専門家ネットワークとは, 科学技術に関する動向や見解などを収集するための, 産学官の研究者, 技術者, マネージャーなど約 2,000 名の専門家集団であり, メンバーは紹介による有意抽出法で選出される。調査時点での所属の構成比率は大学 60.4%, 企業 22.7%, 公的機関・団体 16.0%, その他 0.9%であり, 多分野かつ幅広い年齢層の研究者が含まれること, 企業の研究者が含まれることから調査対象とした。質問は 35 問とし, データ公開経験, 公開データの利用経験, 論文をオープンアクセス (Open Access, OA) とした経験 (OA 誌での出版やセルフアーカイブなど), およびデータ公開の動機や障壁などについて尋ねた。有効回答 1,406 名 (回答率 70.9%) のうち, 現在, 研究活動 (口頭発表や論文出版) を行っている 1,398 名 (70.5%) の回答を対象として, Clarivate Analytics 社の研究動向・統計データベースである Essential Science Indicators (ESI) の 22 分野, 所属機関, 年齢ごとに分析した。

雑誌のデータ公開ポリシーは, ESI の雑誌リストを用いてインパクトファクターの高い原著論文誌を 22 分野から各 10 誌ずつ抽出して調査対象とした。合計 220 誌のウェブサイトでは投稿規定などを調査し, リポジトリにデータを公開して論文に識別子を記すポリシー (Repository Policy, 以下「RP」と, 補足資料にデータを掲載するポリシー (Supplemental Materials' Policy, 以下「SP」) を, 要求の強度, すなわち「必須 (require)」, 「推奨 (recommend)」, 「受諾 (accept)」, 「なし」に従って 4 段階に分類した。また, Ulrichsweb.com などを用いて雑誌の特徴を調査した。その上で, RP の掲載率と日本の研究者によるデータ公開率を分野ごとに比較した。

【結果】

[RQ1]日本の研究者はどの程度データを公開しているのか

日本の研究者を対象とした質問紙調査の結果、データ公開の経験をもつ回答者の割合（以下、「公開率」）は、51.0%であることがわかった。OA の経験をもつ回答者の割合は 70.9%であり、データの公開率の方が約 20 ポイント低かった。公開率を所属機関別にみると、大学（56.9%）、公的機関・団体（55.6%）、企業（32.4%）の順に高かった。

[RQ2]日本の研究者によるデータ公開の内的要因は何か

日本の研究者にとってのデータ公開の動機について、データを公開する理由の 1 位は「研究成果の認知度向上」（58.5%）、2 位は「雑誌のポリシー」（43.8%）であった。複数の先行研究でデータ公開の動機とされていた「分野の規範」は分野別にみても選択率が低く、選択率が高い分野であっても数学（28.6%）、植物・動物学（25.0%）、分子生物学・遺伝学（22.6%）にとどまった。また、科学的利他性に関する項目、すなわち「科学研究や成果実装を推進したいから」（26.1%）、「他の研究者からのリクエストに応じて」（25.4%）、「オープンデータに貢献したいから」（10.5%）も選択率が高いとは言えなかった。

データ公開の障壁について、データの非公開理由の 1 位は「雑誌のポリシーではないこと」（26.4%）、2 位は「時間がないこと」（25.9%）であった。1 位～5 位の選択率はいずれも 20% 台であり（3 位「所属機関にポリシーがないから（22.4%）」、4 位「ニーズがないと思うから」（22.1%）、5 位「業績にならないから」（21.7%））、突出した理由はみられなかった。論文の非公開理由は、1 位「雑誌のポリシーではないこと」（60.4%）と 2 位「公開のための資金」（39.6%）に集中しており、かつ、解決すれば公開して良いと考える回答者が 78.1%にのぼったのに対して、データは 28.4%にとどまった。また、データ公開につながると考えられる非公開理由を特定するために、データ公開意思をもつ回答者群ともたない回答者群とで選択率を比較したが、差のある項目はなかった。ただし、データ公開の業績化や評価、公開データの引用が保証されるならばデータ公開に前向きになるとの意見もみられた。自由記述では、純粋数学など公開するデータがない分野の存在が確認された。

データを管理・公開するための資源については、「人材」、「時間」、「資金」の充足度が低かった。ここでは所属機関による差がみられ、いずれの資源も大学、公的機関・団体、企業の順に充足度が低かった。また、分野別のデータ公開率と公開用リポジトリの充足度には正の相関がみられた。

データを公開する場合の懸念は「引用せずに利用される可能性」と「先に論文を出版される可能性」（以下、「キャリアリスク」）が強かった。「引用せずに利用される可能性」について、自由記述では「盗用」や「剽窃」といった意図的な利用を示唆する表現がみられた。「先に論文を出版される可能性」については、論文の出版前に公開することに対する強い抵抗がうかがえ、「自殺行為」といった表現もみられた。属性別にみると、若年層や大学に所属する研究者は、キャリアリスクに対する懸念が特に強かった。一方、企業の研究者は「機密・プライ

バシー情報」や「商用利用される可能性」への懸念が強く、特許との関わりを指摘する回答も 17 件みられた。分野別にみると、臨床医学、免疫学の研究者は「機密・プライバシー情報」や「商用利用される可能性」への懸念が強かった。

データの公開経験をもつ回答者群は、(a)個人的なデータの提供頻度、(b)個人的なデータの被提供頻度、(c)公開データの入手経験、および利用頻度 ((c-1)自身の研究の参考にする、(c-2)再利用・再分析する、(c-3)結果を再現・追試する)が高く、(d)公開データの利用希望、(e)OA の経験を有する回答者の比率も高いことがわかった。また、データを公開した場合、異分野の研究者による理解が容易であると考えている回答者群は、データ公開率が高かった。一方、データ公開経験と扱っているデータの種類、すなわち、データ量、機密情報の有無、データの所有権の有無には差が認められなかった。

データ公開の経験がない研究者のうち、現在の非公開理由が解決したとしても公開しないと回答した研究者（以下、「データ公開に消極的な研究者」）は、(a)データの提供頻度、(b)被提供頻度、(c)公開データの入手経験や(c-2)再利用・再分析の頻度が低く、(d)公開データの利用を希望する回答者も少なかった。

[RQ3]データ公開の外的要因はどのような状況にあるのか、この外的要因と日本の研究者によるデータ公開状況に乖離はあるのか

学術雑誌のデータ公開ポリシーを調査した結果、調査対象である 220 誌のうち、何らかのデータ公開ポリシーを掲載していた、すなわち RP（リポジトリポリシー）か SP（補足資料ポリシー）のどちらか、あるいは両方を掲載していた雑誌は 204 誌（92.7%）であった。RP の掲載率は 59.5%、SP の掲載率は 89.5%であった。RP は要求の強度が高い「必須」と「推奨」が多く（合計 59.1%）、SP は要求の強度が低い「受諾」が多かった（62.3%）。

分野別に確認すると、生物学と医学の関連領域は RP の掲載率や強度が高い傾向にあり、定評のある生物データのリポジトリへの公開を求めている。しかし、農業科学は 10 誌のうち 7 誌、薬理学・毒性学は 6 誌、精神医学・心理学は 2 誌であるなど、領域内で差がみられた。データ共有による二次分析が盛んな地球科学は 7 誌、宇宙科学は 6 誌、社会科学は 4 誌であり、RP の掲載率は中程度であった。研究内容によってデータが用いられていると推測される分野であっても、コンピュータサイエンスは 2 誌、数学は 1 誌にしか掲載されておらず、工学と経済学・経営学は全く掲載されていなかった。

調査対象誌の特徴として、(1)倫理や利益相反（Conflict of Interest, COI）の開示に関する記述の有無、(2)出版者（学協会と商業出版社）、(3)OA ステータスを調べた。(1)倫理や COI の開示に関する記述がある雑誌は、ポリシーの掲載率や要求の強度が高いことがわかった。(2)出版者は、先行研究では学協会の掲載率が高いという結果が示されていたが、本調査では商業出版社の方が掲載率も要求の強度も高かった。なお、(1)と(2)については分野によるばらつきが大きかったため、生物・医学グループと自然科学グループにわけて確認したが、分野によるばらつきを勘案しても、ほぼ同じ結果であった。(3)OA ステータスについて、OA 誌は全

てポリシーを掲載しており、RP は掲載率も強度も OA 誌、制約つき OA 誌、非 OA 誌の順に高かった。一方、SP については、掲載率は制約つき OA 誌 (92.8%)、非 OA 誌 (82.5%)、OA 誌 (73.3%) の順に高く、「必須」の比率は非 OA 誌 (23.8%)、制約つき OA 誌 (8.8%)、OA 誌 (6.7%) の順であった。オープンサイエンス政策では、研究成果へのパブリックアクセスの観点から論文の OA とデータ公開を推進しているが、雑誌のポリシーについては、OA 誌の方がデータ公開を強く要求しているとまでは言えなかった。

データ公開の要求状況に対して日本の研究者はどの程度データを公開しているのかを確かめるため、分野別に RP の掲載率とデータの公開率を比較した結果、両者には正の相関がみられた。つまり、RP の掲載率が高い分野ほどデータ公開率が高いことがわかった。一方で、RP の掲載率と比較したデータ公開率に乖離がある分野もみられた。RP の掲載率に対してデータ公開率が高く、雑誌のポリシーによらず自発的にデータを公開していると考えられたのは植物・動物学とコンピュータサイエンスであった。この 2 分野に共通していたのは、データの公開理由として「科学研究や成果実装の推進」の選択率が高いこと、データの被提供頻度が高いこと、商用利用される可能性や機密・プライバシー情報への懸念が低いことであった。RP の掲載率に対してデータ公開率が低くデータ公開が困難であると考えられたのは、臨床医学と精神医学・心理学であった。臨床医学はデータの提供頻度や入手経験が低く、再分析・再利用を行う回答者は全くいなかった。データに機密情報を含む回答者の比率が高く、機密・プライバシー情報への懸念も高かった。精神医学・心理学は RP の掲載率が低く、リポジトリの充足度やデータの提供・被提供、入手や利用の頻度も低かった。また、データに機密・プライバシー情報を含む回答者の比率が高かった。ただし、データ公開意思がある回答者の比率が比較的高く、データを利用したいとする他分野の回答者の比率も高かった。

【考察】

日本の研究者によるデータ公開の実践状況、すなわちデータ公開率 (51.0%) は、国外の研究者と比較するとやや低かった。助成機関や研究機関によるデータ公開の義務化がほとんど行われていないことが主な要因の 1 つであると考えられる。

分野別の公開率は、最も高い植物・動物学 (84.2%) から最も低い精神医学・心理学 (30.8%) まで分野間の差がみられた。しかしデータの公開理由のうち「分野・コミュニティの規範」は選択率が低く、上位の分野であっても 3 割に満たなかった。一方、データ公開用リポジトリの充足度については分野による差があり、充足度が高い分野ほど公開率も高かった。

論文はデータよりも公開率が高く (70.9%)、分野による差がみられたが、データ公開との関連はみられなかった。つまり、論文の公開率が高い分野であっても、データの公開率が高いとは限らなかった。OA は雑誌のポリシー (60.4%) と資金 (39.6%) が主な非公開理由であり、解決した場合に公開意思がある回答者は 78.1%にのぼったが、データは 3 割以上の回答者が選択した非公開理由はなく、解決したとしても公開する意思がある回答者は 28.4%であった。データ公開は OA と比較して実践されておらず、成熟度が低いといった指摘を支持

する結果であった。

所属別では、大学（56.9%）、公的機関・団体（55.6%）、企業（32.4%）の順に公開率が高かった。大学の研究者は人材、時間、資金の充足度が低く、キャリアリスクへの懸念が強かった。企業の研究者は機密・プライバシー情報や商業利用される可能性への懸念が強かった。年齢層別では、先行研究と同様に若年層ほどキャリアリスクへの懸念が強かった。

データ公開経験のある回答者群は、経験のない回答者群と比較して、公開用のリポジトリ、時間、人材、資金の充足度が高く、商用利用される可能性や機密・プライバシー情報への懸念は弱かった。また、データの共有や入手、利用の頻度は高く、公開データの利用希望も多かった。データ公開経験のない回答者群のうち、公開意思のある回答者は公開意思のない回答者と比較して、先に論文を出版される可能性や商用利用される可能性への懸念が弱く、データ共有や入手、利用の頻度も高かった。なお、資源の充足度には差がみられなかった。

データ公開に関するさまざまな障壁が明らかになったが、「これを解決すればデータ公開が進む」と判断できるような単一の課題は特定されなかった。つまり、現状ではデータ公開に関する課題は多く、データ公開自体が緒に就いた段階であると考えられる。そこで、ここまでの分析から課題の優先順位を検討するために、(A)データの公開理由のうち選択率の低い項目、(B)データの非公開理由のうち選択率の高い項目、(C)データ公開の障壁のうち選択率の高い項目、(D)データ公開経験がある研究者にとって低い障壁、(E)データ公開に消極的な研究者にとって高い障壁、(F)自発的にデータ公開を行っている分野の研究者にとって低い障壁、(G)データ公開が困難な分野の研究者にとって高い障壁のうち、2項目以上あてはまる要因を主要な課題とみなすこととした。その結果、主要な課題として11の要因が抽出された。学術機関、学術基盤の提供者、学術出版社、助成機関、政府関係者といったステークホルダーが対策を講じる観点から整理して、(1)資源に関する課題（公開用リポジトリ、公開のための時間、人材、資金）、(2)制度に関する課題（業績、先取権）、(3)技術的解決が可能な課題（機密・プライバシー情報、誤解や誤用の可能性）、(4)規範・慣習の醸成が必要な課題（引用、商用利用）に分類した。これらの課題は相互に関連しており、1つの対策が複数の効果やインセンティブとなる可能性や、複数の対策によって1つの懸念を低減する可能性があると考えられる。

また、調査を通じてデータ公開に適さない場合があることが示唆された。分野では、純粋数学のようにデータを扱わない分野、臨床医学、免疫学など機密・プライバシー情報や商業利用への懸念が強い分野、工学のように製品や企業秘密を含む分野などが該当する。また、物性物理学は他の研究者が理解し再利用することが困難であるとの指摘もみられた。企業の研究者は機密・プライバシー情報と商用利用される可能性への懸念が特に強く、特許のため公開は困難であることが明らかになった。データ公開に係る制度設計においては一律の公開を求めないことや、こうした分野の研究者が評価などで不利益を被ることがないように留意する必要があると考えられる。

本研究の主な新規性は、日本の研究者によるデータ公開の実践状況や認識を明らかにして国外の調査との比較を可能にしたこと、データ公開と OA を比較したこと、企業の研究者に関する知見を明らかにしたこと、分野別の雑誌のデータ公開ポリシーと実際のデータ公開率の関連を示したことである。また、COI の開示や倫理に関する記述がある雑誌はポリシーの掲載率や強度が高く、研究不正の抑制に関連がある可能性が示唆された。これらの成果は、オープンサイエンス政策を適切に進めるための議論にも資すると考えられる。

今後、公開データの検索用データベースや引用索引が整備されれば、計量書誌学などの方法を用いてデータ公開の実践状況や分野間の比較が行えるようになると考えられる。また、データ公開を取り巻く情報通信技術や法制度などの環境、および研究者の認識は経年的に変化すると推測される。継続的に、多角的な方法で調査を行うことによって、データ公開による学術情報流通の変容を捉えたい。

Abstract

Open Data Practices and Perceptions of Researchers in Japan

Introduction

In fields such as earth science, astronomy, genomics and particle physics, progress is accomplished by sharing observation data and/or data with a high collection cost. Since around 2000, data-centric e-Science has spread into more fields, and governments and international organizations around the world have developed policies for the disclosure of research data. In addition to improvements in research efficiency and reductions in research and development expenses, there are social aspects to these developments, such as countermeasures against research misconducts, research feedback to citizens, as well as technical aspects such as the development and popularization of information and communication technology and the prosperity of data-intensive science. Indirectly, contributing factors to these developments include the active use of data in society, such as corporate big data collection and the active use of data scientists to analyze those data, the spread of open data by government organizations, and the development of new services which rely on open data. In addition, enabling the utilization of research data beyond subjects and regions is already benefiting society. New scientific discoveries are resulting from the fusion of data from different fields, solutions to social problems are being developed due to the cooperation of industry, academia, government and citizens, economies are expanding as a result of innovation, there has been an expansion of citizen participation in science (Citizen Science), data are being utilized in education, and so on.

Because of this, the policy of open science is expanding globally under the leadership of the Organization for Economic Cooperation and Development (OECD) and the G7 Science Ministers' Meeting. Open science is a policy to make articles and data, resulting from publicly-funded research, freely accessible to citizens. Since 2015, there has been active discussion about open science in Japan. In 2016, the "Fifth Science and Technology Basic Plan" was published, which positioned the promotion of open science as an issue of academic policy.

However, many other problems hamper the development of open data policies, such as the need for infrastructure improvements and the concerns of researchers. It has been pointed out that open science policy should be appropriately promoted based on individual country and subject characteristics. It is also been pointed out that the international competitiveness of researchers in Japan is in decline due to decreased research and development (R&D) expenditure and research time. Therefore, careful consideration is required to promote an open science policy that does not impede research activities. However, at present, the situation and concerns of researchers in Japan regarding open data, and the differing situations and perspectives of different disciplinary fields

have not been sufficiently clarified.

Purpose

The purpose of this study is to clarify open data practice and the perception of open data by researchers in Japan. The researchers' perceptions are revealed by internal factors such as motivation and barriers of data disclosure and experience of using open data. Then, the policy of the scholarly journal is investigated as an external factor which promotes the data disclosure, and it is compared with the current state of the data disclosure by the researchers in Japan. The following three research questions (RQ) were identified: [RQ1] How much data do researchers in Japan disclose; [RQ2] What are the internal factors underlying the release of data by researchers in Japan; [RQ3] What are the external factors of open data, and how do these differ from state of open data in Japan.

Methods

A questionnaire survey was conducted from November to December 2016 to clarify [RQ1] and [RQ2]. In order to clarify [RQ3] and to understand the requirement situation of data disclosure in each field macroscopically, the open data policy of scholarly journals was investigated from April to June 2014. The results of the questionnaire survey and the policy survey of scholarly journals were then compared by subject.

The questionnaire survey was conducted on 1,983 researchers registered in the Network of Science and Technology Experts managed by the National Institute of Science and Technology and Academic Policy (NISTEP) of the Ministry of Education, Culture, Sports, Science and Technology. The Network is a group consisting of approximately 2,000 specialists, such as researchers, engineers, and managers of industry, university and government, who collect trends and opinions on science and technology. Members are selected by the significant extraction method by the introduction. The composition ratio of affiliation, at the time of the investigation, was 60.4% university, 22.7% enterprise, 16.0% public organizations and institutions, and 0.9% other. The Network was considered an appropriate object of research, due to its multidisciplinary membership, wide age range, and its inclusion of enterprise researchers. The questionnaire consisted of 35 questions. Subjects were asked about their experiences with open data, use of public data, experiences with Open Access (OA), such as publishing in OA journals, self-archiving, and so on, and perceived or experienced motivation or barriers to open data. Of 1,406 valid responses (response rate of 70.9%), 1,398 responses (70.5%) were from researchers who make oral presentations or publish articles. Essential Science Indicators (ESI), a database of research trends and statistics published by Clarivate Analytics, was used to analyze these 1,398 responses by subject (22 in total), affiliate organization, and age.

Using the ESI journal list the open data policy of the 10 journals with the highest impact factor for each of the 22 subject areas was examined. The policies of the 220 journals, regarding requirements to publish data in a repository and to write an identifier in the article (Repository Policy, "RP"), as well as policies to post data in supplementary material (Supplementary Materials Policy, "SP"), were classified into 4 levels according to the strength of the request. These levels were "Required", "Recommended", "Accept", "Not mentioned". Journal features were further investigated using Ulrichsweb.com. Finally, the rate of RP and the rate of open data disclosure by researchers in Japan were compared for each subject.

Results

[RQ1] How much data do researchers in Japan disclose?

The results of the questionnaire survey revealed that 51% of respondents have open data experience, and 70.9% have open access experience, and the disclosure rate of data was lower by approximately 20 points. The university disclosure rate was highest (56.9%), followed by public organizations and institutions (55.6%), and enterprises (32.4%).

[RQ2] What are the internal factors behind the release of data by researchers in Japan?

Among the motives and barriers to open data for researchers in Japan, the primary motive for open data was "Raising awareness of research results" (58.5%), followed by "Journal policy" (43.8%). With regard to 'disciplinary practice', even when the selection rate of "discipline in the subject." which was the motive of the disclosure in the preceding research was low even in the viewpoint of the subject independence, and even in the high subject, Mathematics (28.6%), Plant and Animal Science (25.0%), Molecular Biology and Genetics (22.6%). Motivation related to scientific altruism was also not high. For example, "I want to promote scientific research and implementation of results" scored 26.1%, "Upon requests from other researchers" score 25.4%, and "I want to contribute to open data" score only 10.5%.

Regarding the barriers to open data, the top reasons were "Journal has no policy" (26.4%), "Lack of time" (25.9%), "Funding agency does not have a policy (22.4%)", "I think there is no need" (22.1%), and "It is not evaluated as a research achievement" (21.7%), all with similar response levels. Reasons for not publishing OA articles were concentrated in the two highest responses, "Journal has no policy" (60.4%) and "Funds for OA" (39.6%). Further, 78.1% of respondents believed they would be allowed to publish OA articles if they could resolve certain problems, while only 28.4% of the data. In order to identify the reasons for nondisclosure that might lead to open data, reasons for nondisclosure were explored. A difference in the selection rate was found between respondents who intended to disclose data and those who did not. However, some respondents were of the opinion that if the achievement and evaluation of data disclosure and citation for open data are guaranteed, then they would have a positive attitude towards data disclosure. In the 'further

comments' section of the questionnaire, it was confirmed that some fields, such as pure mathematics, do not have open data.

With regard to the availability of resources for managing and disclosing data, satisfaction with "Human resources", "Time", and "Funds" was low. There was a difference according to organization membership, but satisfaction with resources was low in descending order of university, public organizations and institutions, and enterprise. In addition, a positive correlation was found between data disclosure rate by field and adequacy of public repositories.

Concerns over disclosure of data included "The possibility of being used without citation" and "Possibility of publishing an article first" (career risk). Regarding "the possibility of being used without citation" in the 'further comments' section, respondents expressed concern about plagiarism. With regard to "Possibility of publishing an article first", there was strong resistance to opening a manuscript up to the public before publication, with some referring to it as a "suicidal act". Younger age groups and university-affiliated researchers were particularly concerned about the risk to their careers. Meanwhile, researchers in private enterprise exhibited a strong concern for "Confidential and private information" and "Possibility of commercial use", while 17 responses indicated concern about patents. By field, researchers in Clinical Medicine and Immunology showed strong concern regarding "Confidential and private information" and "Possibility of commercial use".

It was found that researchers in Japan with experience of open data have a frequency of personally providing data, a high frequency of personally receiving data, and a high frequency of obtaining and using public data. Researchers revealed that they obtain and use public data as reference for their own research, for re-use and re-analysis, and to reproduce and retest results. The desire to use public data was also high among respondents, as was the proportion of respondents with OA experience. In addition, the data disclosure rate was high for those respondents who thought that researchers in different fields would easily understand disclosed data. On the other hand, no difference was revealed between data disclosure experience and the type of data handled, that is, the amount of data, existence of confidential information, or existence of data ownership.

Among those researchers with no data disclosure experience, those who said they would not disclose data even if current nondisclosure issues were resolved (see below, "Researchers reluctant to release data") showed low frequency of providing data, low frequency of receiving data, low frequency of obtaining data and low frequency of reusing and re-analyzing data. In addition, the number of respondents who desired to use data was low.

[RQ3] What are the external factors of open data, and how do these differ from state of open data in Japan.

Of the 220 journals surveyed, 204 (92.7%) had either RP, SP, or both. The rate of RP was 59.5%, and SP 89.5%. RP had high "Required" and "Recommended" policies (59.1% total), and SP had low "Acceptance" policies (62.3%).

When it was confirmed by subject, the publication rate and rate of RP was high in biology- and medicine-related subject areas, where the public disclosure of public biological data to the repository is required. However, seven out of 10 Agricultural Sciences journals, six out of 10 Pharmacology & Toxicology journals, and two out of 10 Psychiatry/Psychology journals, displayed regional differences. In seven Environment/Ecology journals, six Space Science journals, and four Social Sciences journals secondary analysis by data sharing was popular, and the publication rate of RP was moderate. Even in the subjects where data are supposed to be used according to research content, RP was low, with only two Computer Science journals, one Mathematics journal and no Engineering, Economics and Business journals.

Journals were examined according to the following: (1) existence of descriptions on ethics and disclosure of Conflict of Interest (COI), (2) publisher (academic societies or commercial publishers), (3) OA status. (1) Journals with descriptions of ethics and disclosure of COI showed both higher publication rates of policies and stronger policies. (2) Although previous research has shown that academic societies have higher publication rates, this research found that commercial publishers had higher publication rates and stronger policies. The results of (1) and (2) were almost identical, even when variation by subject was taken into account. However, (3) OA status, showed that all OA journals reported the policy, and RP had high publication rates and strength following the order of OA journals, constrained OA journals, and non-OA journals. At the same time, with regard to SP, the publication rate was highest in constrained OA journals (92.8%), followed by non-OA journals (82.5%), and OA journal (73.3%). The ratio of "Required" was highest in non-OA journals (23.8%), followed by constrained OA journals (8.8%), OA journals (6.7%). According to journals' open science policies, OA and open research data were promoted from the perspective of public access to research results. However, OA journal policy does not make strong demands for open data.

By comparing the publication rate of RP with the publication rate of data by subject, a positive correlation was found in both, which confirmed the degree to which researchers in Japan open their data to open data requests. In other words, the open data rate was high when the publication rate of RP was high. However, in some areas the open data rate was different to the RP publication rate. In Plant and Animal Science and Computer Science the open data rate was higher than the publication rate of RP and data were voluntarily disclosed irrespective of journal policy. There was a high selection rate by respondents from these two subjects for "Promotion of scientific research and implementation of results" as a reason for data disclosure. Respondents from both these subjects show a high frequency of data provision, while the possibility of commercial use and concerns over confidential and private information were low. In Clinical Medicine and Psychiatry/Psychology, the open data rate was low for the publication rate of RP and open data was perceived to be difficult. In Clinical Medicine, the offer frequency and acquisition experience of data were low, and no respondent cited data re-analysis and re-use. The proportion of respondents who included concerns about confidential information in the data was high, and concerns about

confidential private information was also high. In Psychiatry/Psychology, the posting rate of RP was low, and the frequency of offer, acquisition and utilization of the repository was also low. A high proportion of respondents included confidential private information in the data. However, the proportion of respondents who intended to publish open data was comparatively high, and the proportion of respondents in other subject areas, who want to utilize open data, was also high.

Discussion

The state of open data by researchers in Japan (51.0%), is slightly lower than that of foreign researchers. One of the major factors is the lack of mandates for data disclosure by funding agencies and research institutes.

A large difference was revealed between the open data rate of subject areas, from Plant and Animal Science (84.2%) to Psychiatry/Psychology (30.8%). Among the reasons for the disclosure of data, fewer than 30% of respondents selected "discipline and community norms", even among those subjects with high rates of open data use. Differences were revealed according to subject area regarding degree of satisfaction with the repository for open data, and high disclosure rates with high degree of satisfaction.

The rate of OA (70.9%) was higher than the rate of data, and a difference existed by subject area, but no relationship was found with data publication. In other words, even in those subjects where the publication rate is high, the publication rate of data is not always high. Journal policy (60.4%) and funding (39.6%) were the main non-disclosure reasons for OA, and those respondents who intended to disclose when issues were resolved reached 78.1%. Fewer than 30% of respondents chose each non-disclosure reason, and respondents who intended to disclose data even when issues were resolved was 28.4%. Compared to OA, open data was not widely practiced, and the result indicates that maturity of the process is low.

Universities (56.9%) had the highest proportion of open data, followed by public institutions and organizations (55.6%), and enterprises (32.4%). University researchers had lowest levels of human resources, time, and funding, and were strongly concerned about career risks. Enterprise researchers were concerned about sensitive and private information and potential commercial use. As shown in previous studies, younger researchers were more concerned about carrier risk.

Respondents with open data experience had higher levels of repository, time, human resources, and funding for disclosure than those without such experience. Concerns about potential commercial use and sensitive and private information were weak. The frequency of sharing, acquisition and utilization of the data was high, and the utilization of the open data was also high. In addition, among respondents with no open data experience, those with no intention to publish were less concerned about the possibility of publication or commercial use of their articles than those who intended to publish, and they shared, obtained, and used data more frequently. No difference was found in satisfaction with resources.

Although various barriers to open data were identified, no issues were identified that could be judged as "If these barriers were removed, data will be open". In other words, open data currently faces many problems which emerged at the start of the open data era. Therefore, based on the analysis so far, the issues facing open data can be prioritized thus: (A) the low selection rate among reasons for open data, (B) the high selection rate among reasons for open data, (C) the high selection rate among barriers to open data, (D) the low barriers for researchers with open data experience, and (E) the passive barriers to open data. As the result, 11 factors were extracted as priority problems. Stakeholders, such as academic institutions, providers of academic infrastructure, academic publishers, funding agencies, and government officials, are classified into the following categories: (1) problems related to resources (public repositories, time, human resources, funds), (2) problems related to the system (evaluated as research achievement, priority), (3) problems that have technical solutions (confidential and private information, possibility of misinterpretation and misuse), and (4) problems that require the fostering of norms and customs (data citation, commercial use). These issues are interrelated, and one countermeasure is expected to lead to multiple effects and incentives, and multiple countermeasures may reduce one concern.

In addition, the research indicated some subjects for which open data is not suitable or needs to be given special consideration. This applies to subjects that do not handle data, such as pure mathematics; subjects which make use of confidential and private information, such as Clinical Medicine and Immunology; and subjects that have particular concerns about commercial use of data, such as Engineering. It was also pointed out that condensed matter Physics is difficult for researchers in other disciplines to understand and reuse. Researchers from commercial enterprises were especially concerned about confidential private information, the possibility of commercial use, and the difficulty of disclosure due to patenting issues. In designing an appropriate open data system, it is important that uniform disclosure is not required and that researchers in the above fields are not disadvantaged, such as, for example, when undergoing professional evaluation.

This novel research enabled comparison with overseas surveys to clarify the practice and recognition of open data by researchers in Japan. Open data were compared with OA, the knowledge base of researchers working in commercial enterprises was clarified, and the open data policy of journals by subject was compared to the actual open data rate. The publication rate and strength of the policy were found to be high for journals which provide clear information on disclosure, ethics and COI. The possibility of a restrictive form of open data indicated the possibility for discrimination against researchers in some disciplines. These results will contribute to a discussion for the appropriate advancement of open science policy.

In the future, if databases for the retrieval of open data and citation indices are improved, the comparison between practice situation and subjects will be possible with the use of bibliometric methods. It is also assumed that information communication technology and the legal systems which surround open data and researcher recognition will continue to develop. By continuing to

use multiple methods to carry out this investigation, the author hopes to observe the ongoing changes to scholarly communication by means of open data.

目次

1 章 研究背景と目的	1
1.1 データ公開とオープンサイエンス政策	1
1.1.1 データ公開の拡大	1
1.1.2 オープンサイエンス政策によるデータ公開の推進	2
1.1.3 データ公開の課題と日本の研究者の状況	3
1.2 研究目的と研究課題	7
1.3 研究方法	8
1.3.1 日本の研究者を対象とした質問紙調査	8
1.3.2 学術雑誌のデータ公開ポリシー調査	10
1.3.3 比較分析の枠組み	11
1.4 研究の意義	12
1.4.1 学術情報研究への貢献	12
1.4.2 学術政策決定者や学術出版者への示唆	12
1.4.3 研究データ管理での活用	13
1.5 用語の定義と分野分類	15
1.5.1 用語の定義	15
1.5.2 分野の分類	17
1.6 本論文の構成	20
2 章 文献レビュー	21
2.1 データと研究の関わり	21
2.1.1 データの多様性	21
2.1.2 データ共有やデータ保存の取り組み	25
2.1.3 データの増大とデータサイエンスの隆盛	26
2.1.4 データ公開からオープンサイエンスへの展開	27
2.2 データ公開の意義と効果	30
2.3 データ公開を推進する取り組み	35
2.3.1 学術コミュニティの動向	35
2.3.2 学術情報政策の動向	38
2.3.3 図書館等による研究データ管理サービス	43
2.4 データ公開に関する実態調査	45
2.4.1 データ公開に関する質問紙調査	45
2.4.2 データ公開と関連のある経験	49
2.4.3 データ公開の動機と障壁	49

2.5 学術雑誌のデータ公開ポリシー.....	53
2.5.1 データ公開ポリシーの掲載率と強度の分類	53
2.5.2 データ公開ポリシーと関連のある要因	56
2.6 まとめ	57
3 章 日本の研究者によるデータ公開状況と内的要因.....	59
3.1 方法	59
3.1.1 予備調査.....	59
3.1.2 質問の作成方針と項目.....	60
3.1.3 研究課題と質問の設定.....	63
3.1.4 調査対象.....	66
3.1.5 調査方法と期間.....	67
3.1.6 集計・分析方法.....	68
3.1.7 回答率と回答者の属性.....	69
3.2 結果.....	71
3.2.1 データの公開経験.....	71
3.2.2 データの公開理由.....	72
3.2.3 データの非公開理由と公開意思	74
3.2.4 データ公開の障壁の程度.....	76
3.2.5 データの提供と被提供経験.....	79
3.2.6 公開データの入手経験と問題.....	79
3.2.7 データのプロファイル	87
3.2.8 分野による差の比較.....	90
3.2.9 所属機関や年齢層による差の比較	115
3.2.10 データ公開経験の有無による比較	124
3.2.11 データ公開意思の有無による比較.....	137
3.3 考察.....	141
3.3.1 日本の研究者によるデータ公開の状況	142
3.3.2 日本の研究者のデータ公開の障壁	143
3.3.3 データ公開と関連のある経験.....	145
3.3.4 日本の研究者が扱うデータのプロファイル	147
3.3.5 分野による差異.....	147
3.3.6 データ公開と論文の OA の差異.....	150
3.3.7 属性による差異.....	152
3.3.8 データ公開経験の有無による差異	153
3.3.9 データ公開に消極的な研究者の特徴	154
3.3.10 データ公開が適さない場合の検討	154

3.4 まとめ	155
4 章 データ公開の外的要因と日本の研究者の状況	159
4.1 方法	159
4.1.1 予備調査	159
4.1.2 本調査の概要	166
4.1.3 調査対象誌の選定	166
4.1.4 ポリシーの強度と公開先リポジトリ	168
4.1.5 雑誌の特徴と仮説	169
4.1.6 分析方法	171
4.2 結果	171
4.2.1 分野別データ公開ポリシー	171
4.2.2 雑誌の特徴：倫理と COI の開示，出版者，OA ステータス	181
4.2.3 分野別ポリシーとデータ公開経験の比較	191
4.3 考察	192
4.3.1 ポリシーの掲載率と要求の強度	193
4.3.2 雑誌の特徴とポリシーの関連	195
4.3.3 雑誌のポリシーの課題	198
4.3.4 雑誌のポリシーと日本の研究者のデータ公開状況	200
4.3.5 日本におけるデータ公開の課題と展望	203
4.4 まとめ	214
5 章 おわりに	217
5.1 主要な結果と結論	217
5.1.1 研究課題と結果	217
5.1.2 結論	220
5.2 研究の新規性	223
5.3 研究成果の応用	226
5.4 研究の限界	227
5.5 今後の課題	228
謝辞	231
文献リスト	233
全研究業績のリスト	267
付録	275

図目次

図 1-1	研究者からみたデータ公開の構造.....	5
図 1-2	比較分析の枠組み.....	11
図 1-3	データの公開レベル.....	16
図 1-4	論文の構成.....	20
図 2-1	データサイエンスの領域と特性, 実現のためのレイヤー.....	24
図 2-2	DCC Curation Lifecycle Model.....	26
図 3-1	カレントデータの管理・公開資源の充足度 (n=1,396)	77
図 3-2	カレントデータを公開する場合の懸念 (n=1,396)	78
図 3-3	公開データの利用目的と頻度 (n=1,060)	81
図 3-4	カレントデータの量 (n=1,393)	87
図 3-5	カレントデータの理解: 同じ分野と異分野 (n=1,396)	90
図 3-6	分野別データ公開経験 (n=1,384)	91
図 3-7	分野別データ公開の誘因 (規範と業績)	96
図 3-8	分野別「データ公開用リポジトリ」の充足度 (n=1,382)	100
図 3-9	分野別「研究中のストレージ」の充足度 (n=1,382)	101
図 3-10	分野別「商用利用される可能性」への懸念 (n=1,382)	102
図 3-11	分野別「機密・プライバシー情報」への懸念 (n=1,382)	103
図 3-12	分野別「引用せずに利用される可能性」への懸念 (n=1,382)	104
図 3-13	分野別「先に論文を出版される可能性」への懸念 (n=1,382)	105
図 3-14	分野別「研究の誤りを発見される可能性」への懸念 (n=1,382)	106
図 3-15	分野別データ公開の懸念 (機密と商用利用)	107
図 3-16	分野別異分野の研究者によるデータの理解の難しさ (n=1,382)	108
図 3-17	分野別公開データの入手経験とデータ公開経験の関連	110
図 3-18	分野別公開データの再分析・再利用を行う頻度 (n=1,052)	111
図 3-19	所属機関別「データ整備・公開のための人材」の充足度 (n=1,388)	116
図 3-20	所属機関別「データ整備・公開のための時間」の充足度 (n=1,388)	116
図 3-21	所属機関別「データ整備・公開のための資金」の充足度 (n=1,388)	116
図 3-22	所属機関別「商用利用される可能性」への懸念 (n=1,388)	117
図 3-23	所属機関別「機密・プライバシー情報」への懸念 (n=1,388)	117
図 3-24	所属機関別「引用せずに利用される可能性」への懸念 (n=1,388)	118
図 3-25	所属機関別「先に論文を出版される可能性」への懸念 (n=1,388)	118
図 3-26	所属機関別「研究の誤りを発見される可能性」への懸念 (n=1,388)	119
図 3-27	年齢層別「先に論文を出版される可能性」への懸念 (n=1,396)	122
図 3-28	年齢層別「引用せずに利用される可能性」への懸念 (n=1,396)	122
図 3-29	年齢層別データ提供経験 (n=1,397)	123

図 3-30	年齢層別データ被提供経験 (n=1,397)	123
図 3-31	データ公開意思の有無と「引用せずに利用される可能性」への懸念.....	138
図 3-32	データ公開意思の有無と「先に論文を出版される可能性」への懸念.....	138
図 3-33	データ公開意思の有無と「誤解や誤用の可能性」への懸念	139
図 3-34	データ公開意思の有無と「商用利用される可能性」への懸念	139
図 3-35	データ公開意思の有無とデータ提供経験.....	140
図 3-36	データ公開意思の有無とデータ被提供経験.....	140
図 3-37	データ公開意思の有無と公開データの再分析・再利用経験	141
図 4-1	18 分野の分類と雑誌の選択 (予備調査)	162
図 4-2	分野別データ公開ポリシーの強度分布 (22 分野／各 10 誌)	177
図 4-3	分野別補足資料ポリシーの強度分布 (22 分野／各 10 誌)	179
図 4-4	分野別雑誌のデータ公開ポリシーの掲載率とデータ公開率の関連	192
図 4-5	日本におけるデータ公開の主要な課題と対策, 期待される効果.....	213
図 5-1	本研究の比較項目と新たな知見.....	221

表目次

表 1-1	Data Citation Index による分野別データレコード数.....	9
表 1-2	ESI の 22 分類と本研究における日本語名称と略称.....	18
表 1-3	S&EI の分類に基づく本研究の 14 分類と ESI の対応.....	19
表 2-1	データの種類.....	23
表 2-2	オープンサイエンスの定義.....	29
表 2-3	データ公開に関する学術コミュニティの主な動向.....	35
表 2-4	データ公開・オープンサイエンス政策（国際）.....	38
表 2-5	データ公開・オープンサイエンス政策に関する主要な文書（日本）.....	41
表 2-6	データ公開に関する質問紙調査の概要.....	46
表 2-7	質問紙調査によるデータ公開の主な動機と障壁.....	50
表 2-8	研究者からみたデータ公開の動機と障壁.....	51
表 2-9	先行研究による学術雑誌のデータ公開ポリシーの掲載率.....	53
表 2-10	先行研究によるデータ公開ポリシーの強度の分類.....	55
表 2-11	ポリシーの掲載率や強度と関連がある雑誌の特徴.....	57
表 3-1	質問項目一覧.....	62
表 3-2	質問の構成.....	63
表 3-3	回答者の年齢層と所属.....	69
表 3-4	回答者の分野と所属.....	70
表 3-5	データ公開経験の有無.....	71
表 3-6	データを公開した方法（n=713, 複数回答）.....	72
表 3-7	データ公開と OA の経験の有無.....	72
表 3-8	データ公開と OA の理由（複数回答）.....	73
表 3-9	データと論文の公開理由の自由記述.....	74
表 3-10	データと論文の非公開理由（複数回答）.....	75
表 3-11	データ公開と OA の意思.....	76
表 3-12	データの提供・被提供状況とデータ公開率.....	79
表 3-13	公開データの入手経験.....	80
表 3-14	公開データの入手方法（n=1,060, 複数回答）.....	80
表 3-15	公開データ入手における問題の有無.....	81
表 3-16	公開データを入手する際の問題（n=846, 複数回答）.....	82
表 3-17	データと論文の探索方法（複数回答）.....	83
表 3-18	データと論文の信頼性の判断基準（複数回答）.....	84
表 3-19	公開データを利用してみたい分野（複数回答）（単位：%）.....	86
表 3-20	カレントデータの所有権をもつ人・組織の有無.....	88
表 3-21	カレントデータの所有権をもつ人・組織（n=1,166, 複数回答）.....	88

表 3-22	カレントデータの機密情報の有無.....	89
表 3-23	カレントデータに含まれる機密情報 (n=683, 複数回答)	89
表 3-24	分野別データ公開率および関連する結果	92
表 3-25	分野別データの公開理由.....	94
表 3-26	分野別データの非公開理由.....	97
表 3-27	分野別データ公開意思.....	99
表 3-28	分野別データ提供・被提供経験.....	109
表 3-29	分野別リポジトリの利用状況.....	112
表 3-30	分野別カレントデータの量.....	114
表 3-31	所属機関別データ公開経験.....	115
表 3-32	所属機関別カレントデータの所有権をもつ人・組織	119
表 3-33	所属機関別機密情報の有無.....	120
表 3-34	年齢層別データ公開経験.....	121
表 3-35	データの整備・公開のための人材の充足度とデータ公開経験.....	124
表 3-36	データの整備・公開のための時間の充足度とデータ公開経験.....	125
表 3-37	データの整備・公開のための資金の充足度とデータ公開経験.....	125
表 3-38	データ保存用ストレージの充足度とデータ公開経験	126
表 3-39	データ公開用リポジトリの充足度とデータ公開経験	126
表 3-40	研究中のデータ用ストレージの充足度とデータ公開経験	127
表 3-41	引用せずに利用される可能性とデータ公開経験	127
表 3-42	先に論文を出版される可能性とデータ公開経験	128
表 3-43	誤解や誤用の可能性とデータ公開経験.....	128
表 3-44	機密・プライバシー情報の懸念とデータ公開経験	129
表 3-45	商用利用される可能性とデータ公開経験	130
表 3-46	研究の誤りを発見される可能性とデータ公開経験	130
表 3-47	データ提供経験とデータ公開経験.....	131
表 3-48	データ被提供経験とデータ公開経験.....	131
表 3-49	公開データの入手経験とデータ公開経験.....	132
表 3-50	データの利用状況とデータ公開経験.....	132
表 3-51	公開データの利用希望とデータ公開経験.....	133
表 3-52	カレントデータの量とデータ公開経験の有無.....	133
表 3-53	カレントデータの所有権とデータ公開経験の有無	134
表 3-54	カレントデータの機密情報とデータ公開経験.....	134
表 3-55	同じ分野の研究者によるカレントデータの理解とデータ公開経験	135
表 3-56	異分野の研究者によるカレントデータの理解とデータ公開経験	135
表 3-57	データ公開経験の有無と関連のある項目	136
表 3-58	データ公開に影響を与える要因.....	137
表 3-59	データ公開意思の有無とデータ入手経験の有無	140

表 3-60	データ公開意思の有無と公開データの利用希望の有無	141
表 3-61	質問項目と分野・属性・データ公開経験・データ公開意思との関連.....	157
表 4-1	データ公開ポリシーの強度の分類（予備調査）	163
表 4-2	分野別学術雑誌のデータ公開ポリシー：予備調査（n=180）	164
表 4-3	出版者別タイトル数（n=220）	167
表 4-4	出版国別タイトル数（n=220）	168
表 4-5	データ公開ポリシーの強度の分類.....	169
表 4-6	データ公開ポリシーの掲載状況.....	172
表 4-7	分野別学術雑誌のデータ公開ポリシー（n=220）	173
表 4-8	分野別 IF（n=220）	175
表 4-9	分野別雑誌年齢（n=220）	176
表 4-10	データの公開先として例示されたりポジトリ（上位 10 件）	178
表 4-11	分野別調査対象誌の特徴（n=220）	182
表 4-12	分野グループ	183
表 4-13	倫理・COI に関する記述の有無とデータ公開ポリシーの有無	183
表 4-14	倫理・COI に関する記述の有無とデータ公開ポリシーの掲載率	184
表 4-15	倫理・COI に関する記述の有無とデータ公開ポリシーの強度	185
表 4-16	倫理・COI に関する記述の有無とデータ公開ポリシー（分野グループ別）	186
表 4-17	学協会・商業出版社別データ公開ポリシーの有無	187
表 4-18	学協会・商業出版社別データ公開ポリシーの掲載率	188
表 4-19	学協会・商業出版社別データ公開ポリシー	188
表 4-20	学協会・商業出版社別データ公開ポリシー（分野グループ別）	189
表 4-21	OA ステータスとデータ公開ポリシーの有無.....	190
表 4-22	OA ステータスとデータ公開ポリシーの掲載率.....	190
表 4-23	OA ステータスとデータ公開ポリシーの強度.....	191
表 4-24	分野別データリポジトリ数（Databib）	194
表 4-25	分野による差がみられた項目と特徴的な分野の結果	201
表 4-26	データ公開の主要な動機と障壁.....	204
表 4-27	日本におけるデータ公開の主要な課題.....	205
表 4-28	リポジトリの利用状況.....	206
表 4-29	データ公開経験・態度と関連がみられた経験・認識	212
表 5-1	本研究の新規性と先行研究の状況.....	224

付表目次

付表 1	先行研究によるデータ公開の動機と障壁の出典.....	275
付表 2	回答者の研究分野	288
付表 3	オープンアクセスの状況.....	288
付表 4	論文を OA にした理由 (n=991, 複数回答)	289
付表 5	論文が OA ではない理由 (n=407, 複数回答)	289
付表 6	オープンアクセスの意思.....	290
付表 7	研究に利用したことがある分野 (n=1,398, 複数回答)	290
付表 8	論文の探索ツール (n=1,398, 複数回答)	291
付表 9	論文の信頼性の判断基準 (n=1,398, 複数回答)	291
付表 10	データ提供経験	292
付表 11	データ被提供経験	292
付表 12	公開データの入手経験	293
付表 13	公開データの入手方法 (n=1,060, 複数回答)	293
付表 14	公開データ入手における問題の有無.....	294
付表 15	公開データを入手する際の障壁 (n=846, 複数回答)	294
付表 16	公開データの利用目的 (n=1,060)	294
付表 17	公開データを研究に利用した経験の有無.....	295
付表 18	研究に利用したことがある公開データの分野 (n=1,034, 複数回答)	295
付表 19	公開データの探索ツール (n=1,034, 複数回答)	296
付表 20	公開データの信頼性の判断基準 (n=1,034, 複数回答)	296
付表 21	利用したい公開データの有無.....	297
付表 22	今後, 利用してみたいと思う公開データの分野 (n=1,357, 複数回答)	297
付表 23	データ公開経験の有無	298
付表 24	データを公開した方法 (n=713, 複数回答)	298
付表 25	データを公開した理由 (n=713, 複数回答)	299
付表 26	データを公開していない理由 (n=683, 複数回答)	299
付表 27	データ公開意思	300
付表 28	データ保存期間規定の有無.....	300
付表 29	カレントデータの量	300
付表 30	カレントデータの所有権をもつ人・組織の有無.....	301
付表 31	カレントデータの所有権をもつ人・組織 (n=1,166, 複数回答)	301
付表 32	カレントデータの機密情報の有無.....	301
付表 33	カレントデータに含まれる機密情報 (n=683, 複数回答)	301
付表 34	カレントデータの望ましい保存期間.....	302
付表 35	カレントデータ公開の有無.....	303

付表 36	カレントデータの管理資源の充実度	303
付表 37	カレントデータを公開する際の障壁	304
付表 38	カレントデータの理解：同じ分野	304
付表 39	カレントデータの理解：他分野	305
付表 40	カレントデータのメトリクスへの関心	305
付表 41	関心のあるメトリクス (n=1,198, 複数回答)	305
付表 42	カレントデータの管理方法への関心	306
付表 43	データの整備・公開に関して詳しく知りたい項目 (n=1,257, 複数回答) ..	306
付表 44	第三者によるカレントデータの管理における専門性	307
付表 45	カレントデータの管理に専門性を必要とする項目 (n=1,302, 複数回答) ..	307
付表 46	回答者の分野と年齢	308
付表 47	分野別同じ分野の研究者によるデータの理解の難しさ (n=1,382)	309
付表 48	所属機関別「誤解や誤用の可能性」への懸念	310
付表 49	所属機関別カレントデータの理解（自分野）	310
付表 50	所属機関別カレントデータの理解（他分野）	311
付表 51	国別データリポジトリ数	311
付表 52	IF の分野内順位とポリシーの強度 (n=220)	312
付表 53	刊行開始年代とポリシーの強度	313
付表 54	雑誌年齢の分野内順位とポリシーの強度 (n=220)	314

1 章 研究背景と目的

研究の成果である論文やデータをインターネットで公開し、学術関係者のみならず、企業や市民による利活用を可能にするオープンサイエンス¹⁾政策が国内外で推進されている。オープンサイエンスの実現によって学術の発展や透明性の向上、イノベーションの創出が期待されており、学術雑誌や学術機関によるデータ公開要求も広がりつつある。

一方、データ公開を実現するためには、インフラの整備や研究者の懸念など多くの課題が存在する。また、オープンサイエンス政策は国や分野による特性をふまえて適切に進める必要があると指摘されている。日本の研究者は研究開発費や研究時間の減少によって、国際競争力が低下していると指摘されていることから、オープンサイエンス政策の推進にあたっては、研究活動を阻害しないよう慎重な検討が必要であると考えられる。しかし現状では、データ公開に関する課題や分野による差異、日本の研究者によるデータ公開の状況は十分に明らかにされているとは言い難い。そこで本研究は、日本の研究者によるデータ公開の実践状況およびデータ公開に対する認識を明らかにすることを目的とする。

本章では、こうした研究背景に基づく研究目的と研究課題を示した上で、研究方法、研究の意義、応用可能性について述べる。また、本研究で用いる用語の定義と分野分類、論文全体の構成を示す。

1.1 データ公開とオープンサイエンス政策

1.1.1 データ公開の拡大

1440 年代半ばごろ Johannes Gutenberg によって活版印刷技術が開発された。それから 220 年を経た 1665 年、印刷技術を用いて世界最初の学術雑誌として *Journal des Sçavans* (Académie des Sciences) と *Philosophical Transactions* (Royal Society) が刊行され、研究の成果は論文の形で学術コミュニティに広く共有されるようになる (King, McDonald, & Roderer, 1981; 逸村, 2015)。1990 年には Tim Berners-Lee によって World Wide Web (WWW) が開発された。そし

¹⁾ オープンサイエンスの正式な定義はまだないと言われているが (OECD, 2015; 国際的動向を踏まえたオープンサイエンスに関する検討会, 2015)、オープンサイエンスに関する政策文書は、おおむね論文のオープンアクセス (Open Access, OA) とデータ公開を対象としている。たとえば G7 科学技術大臣会合は、“幅広い分野の公的資金による研究成果 (論文や関連するデータセット等) に学術関係者だけでなく、民間企業や一般市民が、広く利用・アクセスできるようにするものである” (G7 Science and Technology Ministers, 2016, p. 9) としている。

1 章 研究背景と目的

て現在、インターネットを通じて学術雑誌に掲載された論文などの出版物に、研究者のみならず誰もが自由にアクセスできるオープンアクセス（Open Access, OA）の取り組みが広がっている（倉田, 2007; Suber, 2012; 佐藤, 2018）。

論文と同様に研究の成果であり、論文の根拠となるデータ（以下、「データ」）も同じ領域の研究者の間で共有されてきた。たとえば社会科学分野では、研究者同士で共有したデータを用いた二次分析が行われてきた（Hyman, 1972; Fienberg, Martin, & Straf, 1985; Collins, 1992）。化学や物理学、天文学などの分野では、論文から数量的データを抽出し、評価して *Landolt-Börnstein Tabellen*（1883 年創刊）や *Tables Annuelles de Constantes et Données Numériques*（1910 年創刊）のようなハンドブックとして出版する取り組みも行われてきた（小谷, 1974）。データ量の増加に応じてデータセンターが設立され、地球科学（Diepenbroek, et al., 2002）、天文学（Henneken, 2015）、生命科学（Kaye, Heeney, Hawkins, Vries, & Boddington, 2009）などデータの採取コストが高い分野では世界規模でのデータ共有が行われるようになった。情報通信技術（ICT）の発達やサイバーインフラストラクチャの普及によってデータの共有や公開が容易になり、2000 年ごろからはデータを中心とした科学研究（data intensive science, data driven science）や e-Science の潮流も高まり、より多くの分野のデータがインターネット上で公開されるようになった（McCain, 2000; Edwards, Jackson, Bowkerm, & Knobel, 2007; Akins et al., 2003; Hey, Tansley, & Tolle, 2009; Critchlow & Kleese, 2013; Assante, Candela, Castelli, Manghi, & Pagano, 2015）。

1.1.2 オープンサイエンス政策によるデータ公開の推進

データ公開が拡大する一方で、公開したデータの知的財産権の主張や有償化といった動きもみられるようになった（National Research Council [NRC], 1997）。しかし、研究者や市民ができる限り制限なく公開データを利用することができるようになれば、(1)研究の効率化、(2)異分野データの活用と共同研究の推進、(3)研究の透明性の向上、(4)データの長期保存、(5)市民への研究成果の還元、(6)イノベーションの創出と経済効果、(7)市民科学の進展などが可能になると考えられている（詳細は 2.2 で述べる）。そのため国際組織や各国・地域の政府は、研究データへのパブリックアクセスを求めるようになった。そして、既に取り組みが進んでいた論文や出版物の OA を含めたオープンサイエンス政策が拡がりつつある。

具体例として、主要な政策を挙げる。2007 年に経済協力開発機構（Organization for Economic Co-operation and Development, OECD）は「公的資金による研究データへのアクセスに関する OECD 原則とガイドライン（*OECD principles and guidelines for access to research data from public funding*）」（以下、「OECD 原則」）（Organization for Economic Co-operation and Development [OECD], 2007）を、2013 年に G8 科学大臣会合は科学研究データ公開の基本原則を含む共同声明（G8 Science Ministers, 2013）を公開した。これらを契機として、各国・地域の政府組織や助成機関、学術雑誌、研究機関、学会などによる研究データ公開の取り組みが加速し、義務化の動きも

拡がっている（OECD, 2015）。2016 年に行われた G7 科学技術大臣会合による *Tsukuba Communiqué*（共同声明）では、オープンサイエンスが分野横断的な課題として位置づけられ、“政府機関やその他機関が、データ収集、解析、保存、公表のための適切なインフラとサービスに継続的に投資を行うことが必須である”（G7 Science and Technology Ministers, 2016, p. 9）と記された。

日本の学術情報政策の場においても、G8 科学大臣会合の共同声明を契機としてデータ公開に関する議論が展開され、2015 年ごろからは論文の OA とあわせたオープンサイエンス政策として検討が進められてきた。第 5 期科学技術基本計画（2016～2020 年度）では、“国は、資金配分機関、大学等の研究機関、研究者と連携し、オープンサイエンスの推進体制を構築する”（内閣府, 2016, p. 32）と述べられている。そのフォローアップとして 2018 年 6 月に閣議決定された「統合イノベーション戦略」では、「第 2 章：知の源泉」に、「(2)オープンサイエンスのためのデータ基盤の整備」が掲げられ、“公的資金による研究成果としての研究データについては、データインフラを通して機械判読可能化を促進する”（内閣府, 2018, p. 16–19）ことが主要施策の 1 つとされている。

1.1.3 データ公開の課題と日本の研究者の状況

データ公開は特定の分野のみならず、オープンサイエンス政策によって多くの分野に拡大されつつある。日本においてもオープンサイエンス政策によってデータ公開の要求が具体化されようとしている。一方、データ公開を実現しようとする際には、さまざまな課題があることが指摘されてきた。ここでは本研究の問題意識として、まず、研究者の立場からみたデータ公開の構造と課題を整理する。続いて、データ公開は分野による特性に配慮する必要があると指摘されているにもかかわらず、具体的にどのような問題があるのかが十分に明らかにされていないことを示す。最後に、日本の研究者が置かれている厳しい研究環境について述べる。

[1] 研究者からみたデータ公開の構造と課題

研究者は、データ公開政策などの制度や、業績などの誘因を動機としてデータ公開に至ることが指摘されてきた（Nelson, 2009; Edwards, 2016; Open Research Data Task Force, 2017）。また、公開データの利用経験は、研究者のデータ公開行動に影響を与えているとする研究もみられる（Piwowar, 2011）。しかし実際にデータを公開する際には、公開のための資源の不足感や、データ公開に対する懸念といった障壁が存在するため、データ公開を阻害しているということが指摘されてきた（Nelson, 2009; Royal Society, 2012; OECD 2015; Edwards, 2016）。次章で詳しく述べるように、先行研究によって示された動機や障壁は多岐にわたり、かつ、研究によって調査項目や重要とされる項目が異なるため、どの動機や障壁が日本の研

1 章 研究背景と目的

究者にとって重要であるのかを判断することは難しい。そこで、Fecher, Friesike, and Hebing (2015) と Kim and Stanton (2015) の枠組みを参考としながら、複数の先行研究で指摘されていた動機や障壁、データ公開に関連があると予想される経験を整理して、これを本研究の前提とすることとした(先行研究の詳細は次章で述べる)。ここでは、まず Fecher et al. (2015) と Kim and Stanton (2015) による研究の概要について述べる。

Fecher et al. (2015) は、研究データ共有プロセスの概念枠組みを構築するために、98 報の学術論文のシステマティックレビュー、およびデータの二次利用者 603 名を対象とした調査を行った。その結果、6つのカテゴリ、すなわちデータの提供者、研究機関、研究コミュニティ、規範、データのインフラストラクチャー、データの利用者による概念枠組みを提示した。そしてデータ共有のより良いインセンティブとなるデータ公開ポリシーが、研究成果の質の向上や科学の進展に必要であると指摘した。

Kim and Stanton (2015) は、研究者のデータ共有行動に影響を与える要因を、機関による環境と個人的な動機にわけてモデル化し、それぞれの影響の強さを明らかにするために質問紙調査を行った。調査は 2012 年から 13 年にかけて実施され、STEM (科学・技術・工学・数学)²⁾の 43 分野 1,317 名から得た回答についてマルチレベル分析を行った。その結果、データ共有行動と関連がみられたのは、分野レベルでの規範や雑誌の出版者によるプレッシャー、および個人的なレベルでの科学的利他性 (scholarly altruism)³⁾やキャリアベネフィット (career benefit)⁴⁾に関する認識であった。また、データ共有に労力と時間が必要であることを認識している研究者はデータを共有する可能性が低いという結果を示した。

Fecher et al. (2015) と Kim and Stanton (2015) の研究結果ふまえて、「研究者からみたデータ公開の構造」を作成した (図 1-1)。前述の通り、研究者のデータ共有経験や公開データの利用経験は、データ公開と関連がある可能性が示唆されている (Piwowar, 2011)。また、キャリアベネフィットのような誘因や、学術雑誌のポリシーなどの制度は、データ公開の動機となっていることが指摘されている。しかしデータを公開しようとする際には、資源の不足感やデータ公開への懸念が障壁となっているという調査結果もみられる。図 1-1 の矢印は、データ公開の実現によってさまざまな意義や効果が生まれると期待されているが、障壁があるために一部のデータしか公開に至っていない状況を示している。本研究は、図 1-1 の研究者からみたデータ公開の構造を前提として調査や分析を行う。

²⁾ Science, Technology, Engineering and Mathematics (STEM)

³⁾ 科学的利他性について、Kim and Stanton (2015) は先行研究を参照した上で、他の研究者の研究への援助、オープンな科学研究の支援、科学研究への貢献や推進などを挙げている。

⁴⁾ キャリアベネフィットについては、引用などクレジットの獲得、認知向上、ステータスの向上、キャリアに役立つことなどを挙げている。

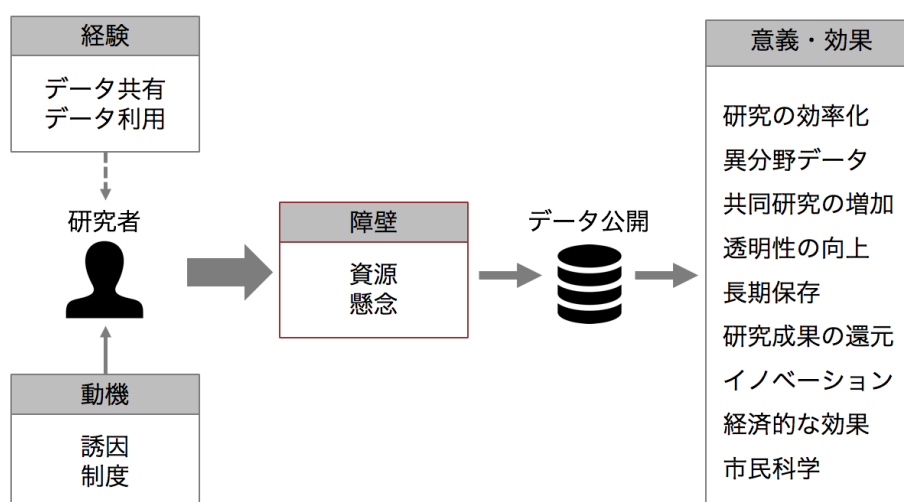


図 1-1 研究者からみたデータ公開の構造

[2] 分野による特性

データ公開に関する政策は OA と比較して成熟度が低く、課題が多いと指摘されている (OECD, 2015)。国内外の政策文書に共通して挙げられている課題は、データ公開やオープンサイエンスの推進にあたって、研究者（個人）の認識のみならず、分野による違いや特性に配慮するべきであるということである。たとえば「我が国におけるオープンサイエンス推進のあり方について：サイエンスの新たな飛躍の時代の幕開け（以下、「内閣府報告書）」では、“各省庁等のステークホルダーは、オープンサイエンスを推進すべき領域、プロジェクトを選定し、研究活動上の利益・損失や研究途上の取扱及び機微の判断など各分野の専門家・研究者、技術者の意見を十分に取り入れ、その分野の活動・研究成果が最大化されることを旨として、オープンサイエンス実施方針を定める”（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015, p.14）としている。「科学技術イノベーション総合戦略 2017」⁵⁾では、“各々の研究分野や研究コミュニティの特性に応じたオープンサイエンス推進のための共通課題等を検討するための検討体制を構築する”（内閣府, 2017, p. 89）と記されている。そして日本学術会議オープンサイエンスの取組に関する検討委員会（2016）は、研究データ同盟（Research Data Alliance, RDA）⁶⁾ Europe（2014）の報告書を参照しながら、“欧州における専門家レポートでは、オープンサイエンスを推進することが重要としながらも、研究データ問題は特に複

⁵⁾ 「第 4 章 科学技術イノベーションの基盤的な力の強化」の「(2) 知の基盤の強化」に「③オープンサイエンスの推進」が示されている。

⁶⁾ RDA は、欧州委員会（European Commission, EC）や米国国立科学財団（National Science Foundation, NSF）、米国国立標準技術研究所（National Institute of Standards and Technology, NIST）などを中心として 2013 年に活動を開始した。目的は、研究者やイノベーターが社会の大きな課題を解決するために技術や学問領域や国を超えてオープンにデータを共有することである。

1 章 研究背景と目的

雑であるため、理解できていない状態で拙速にルール化をすべきではない、という提言が含まれていることにも留意が必要である” (p. 10) と指摘している。

これまで、分野によってデータ公開の状況が異なることや (Womack, 2015), データは天文学やゲノムでしか有効に活用できる状態で公開されていないことが指摘されてきた (Borgman, 2012)。しかし、データ公開があまり行われていない分野やデータ公開が適さない理由、分野によって扱うデータの量や機密情報などに差異があるのかどうかといったことについては、十分に明らかにされているとは言い難い。つまり政策文書で指摘されているような、分野による違いや、どのような点に配慮すればよいのかは明確にされていないと考えられる。

[3] 日本の研究環境

近年、日本の研究環境の問題として、研究者が研究に割ける時間が減少していること (科学技術政策研究所, 2016), および研究開発費や論文の生産性が低迷していることが指摘されている (Phillips, 2017)。Nature Index 2017 Japan⁷⁾によれば、世界の論文数は 2005 年から 2015 年にかけて 80%増加しているのに対して日本の論文数は 14%しか増加しておらず、世界シェアは 10 年で 3 分の 1 に減少している。また、高品質な論文 (high-quality papers) の数が減少していること、天文学を除く 13 分野は論文数の伸び率が世界の伸び率と比較して鈍いことなども明らかにされている。こうした状況を受けて、第 5 期科学技術基本計画では、日本の論文数を増やすこと、および日本の論文の 10%が世界の被引用回数トップ 10%論文になることを目指すとしている (内閣府, 2016, p. 30)。

一方、学術雑誌によるデータ公開ポリシーは経年的に増加している (Stodden, Guo, & Ma, 2013)。たとえば、2016 年 12 月に Springer Nature 社は同社の傘下にある出版社の 600 以上の雑誌に 4 種類のデータ公開ポリシー⁸⁾のいずれかを分野横断で採用すると発表した (Springer Nature, 2016)。また、データを公開している論文は、公開していない論文と比較して被引用回数が多いという複数の調査結果がある (Piwowar, Day, & Fridsma, 2007; Vandewalle, 2012; Henneken & Accomazzi, 2012; Dorch, Drachen, & Ellegaard, 2015)。従って、日本の研究者が学術雑誌のデータ公開ポリシーに対応できない場合は、論文投稿の障壁となったり、研究のビジビリティ向上の妨げになる可能性があると考えられる。また、学術雑誌のデータ公開ポリシーに対応できない場合、1.1.2 に示した利点も逸することになる。たとえば、国際共同研究の機会損失、研究開発費をかけたデータの亡失といった事態が起こりうるのではないだろうか。

Wiley 社の国際調査によって、データ公開の状況は、国によって差があることが明らかにされた (Wiley, 2014)。また、Tenopir et al. (2011) や Schmidt, Gemeinholzer, and Birgit (2016a)

⁷⁾ Nature Index 2017 Japan. <https://www.nature.com/collections/hmjgglbjn>, (accessed 2018-10-31).

⁸⁾ “Research data policies”. Springer Nature. <http://www.springernature.com/gp/group/data-policy/>, (accessed 2018-10-31).

による国際調査でも、データ公開の状況は地域によって差があることが示されている。Schmidt et al. (2016a) の調査を紹介した近藤 (2016) は、日本やアジアにおいて“オープンリサーチデータを推進する際には、国際動向をふまつつも、国・地域の事情に即した施策が必要” (p. 257) であると指摘している。しかし、日本の研究者を対象とした実態調査は、ほとんど行われていない⁹⁾。また、調査時点では日本の主要な助成機関によるデータ公開の義務化が行われていなかったため¹⁰⁾、助成機関の要求が強い公開要因となっている国外の調査とは結果が異なると考えられる。

1.2 研究目的と研究課題

オープンサイエンス政策等によって、データ公開は多くの分野へ拡大されようとしている。その際には分野による違いへの理解と対応が重要であると指摘されているにもかかわらず、データ公開に関する分野横断の比較研究はほとんど行われていない。そもそも、あらゆる分野や状況においてデータ公開が必要であるのか、そうでない場合があるとすれば、それはどのような条件によるものかということについては、十分に検討されているとは言い難い。また、日本においてもオープンサイエンス政策が進められようとしているが、現状では日本の研究者によるデータ公開の実践状況や認識、データ公開に関する分野による差異は十分に明らかにされていない。

そこで本研究は、日本の研究者によるデータ公開状況およびデータ公開に対する認識を明らかにすることを目的とする。この目的を達成するために、まず、日本の研究者はどの程度データを公開しているのかを明らかにする。続いてデータ公開に対する研究者の認識を、図 1-1 (研究者からみたデータ公開の構造) に示したデータ公開の動機や障壁、公開データの利用経験といった内的要因から明らかにする。そして日本の研究者によるデータ公開の状況を相対的に把握するために、データ公開を推進するとされている学術雑誌のポリシーを外的要因として調査し、日本の研究者によるデータ公開の状況と分野別に比較する。次節では、3つの研究課題 (Research Questions, RQ) を明らかにするための研究方法について述べる。

[RQ1]日本の研究者はどの程度データを公開しているのか

[RQ2]日本の研究者によるデータ公開の内的要因は何か

[RQ3]データ公開の外的要因はどのような状況にあるのか、また、この外的要因と日本の研究者によるデータ公開状況に乖離はあるのか

⁹⁾ 国際調査に日本の回答者が含まれている場合もあるが、日本の研究者を対象とした調査は小野、小池、柴崎 (2016) によるもの (回答者 38 名) のみであった。詳細は 2.4 で述べる。

¹⁰⁾ たとえば 2018 年 10 月現在も、科学研究費助成事業 (日本学術振興会) はデータ公開を要求していない。

1.3 研究方法

[RQ1]と[RQ2], すなわち日本の研究者によるデータ公開の状況と内的要因を明らかにするために, 2016 年 11 月から 12 月にかけて質問紙調査を実施した。また, [RQ3]データ公開の外的要因の状況を明らかにするために, 2014 年 4 月から 6 月にかけて学術雑誌のデータ公開ポリシーを調査した。その上で, 質問紙調査の結果(日本の研究者によるデータ公開率)と, 学術雑誌のポリシー調査の結果(データ公開の要求状況)を分野別に比較した。以下では, 日本の研究者を対象として質問紙調査を行った理由, 学術雑誌のポリシーを調査した理由, および比較分析の枠組みについて述べる。

1.3.1 日本の研究者を対象とした質問紙調査

日本の研究者を対象とした調査を行う理由は, データ公開の問題は国や分野によって異なると指摘されているにもかかわらず, 日本を対象とした研究はほとんど行われていないためである。また, 1.1.3 で述べたように日本の研究環境は厳しい状況にありながら, 論文の生産性や被引用回数を向上させることが求められている。そのためには, 学術雑誌のデータ公開ポリシーに対応する必要があると考えられるが, 現状ではどの程度データ公開が要求されているのか, それに対してデータ公開を実践できているのか, 課題はあるのか, あるとすればどのような課題なのかといったことが十分に明らかにされていない。そこで本研究は日本の研究者が扱っているデータのプロフィールや公開状況, データ公開に対する認識を調査し, 学術雑誌のポリシーによるデータ公開の要求状況や国外の調査結果と比較分析するとともに分野別の課題を明らかにすることとした。

質問紙調査を行う理由は, データベースなどを用いた調査が困難なことである。OA 論文についてはデータベースを用いた多数の調査が行われている。たとえば Archambault et al. (2013) は, Directory of Open Access Journals (DOAJ)¹¹⁾, PubMed Central¹²⁾, Scopus¹³⁾を用いて 2008 年から 2011 年に出版された査読付き論文 16 万報を対象として, 22 分野の OA 比率を算出した。その結果, 科学技術一般の比率が最も高く (64%), ヴィジュアル・パフォーマンス・ツールが最も低い (13%) ことを示している。これに対してデータの引用索引データベースには Clarivate Analytics 社¹⁴⁾の Data Citation Index (DCI) があり, 分野ごとのデータレコード数を把握することができる (Ikeuchi & Itsumura, 2015)。しかし, 本研究を行った 2016 年の段階では

¹¹⁾ Directory of Open Access Journals (DOAJ) <https://doaj.org/>, (accessed 2018-10-31).

¹²⁾ 米国生物工学情報センター (National Center for Biotechnology Information, NCBI) の生物医学・生命科学論文アーカイブ。現在の名称は PMC。 <https://www.ncbi.nlm.nih.gov/pmc/>, (accessed 2018-10-31).

¹³⁾ Elsevier 社の抄録・引用文献データベース。 <https://www.scopus.com/>, (accessed 2018-10-31).

¹⁴⁾ 2016 年 10 月までは Thomson Reuters 社。

収録レコード数が増減しており不安定であった。つまり、調査時点によって結果が大きく変わると考えられる。たとえば表 1-1 に示すように、化学のレコード数は 2015 年の 42,207 件から 2016 年の 144,316 件へと、他の分野と比べて急激に増えている。一方、レコード数が減少している分野もみられる¹⁵⁾。たとえば、人口統計学は 2014 年の時点では 41,854 件であったが 2016 年には 5,458 件となっている。また、1 レコードあたりのデータやデータセット数はリポジトリによって異なるため、レコード数をもって分野間のデータ公開状況を比較することは難しいと考えられる。日本のレコード数についても、海洋学は 2015 年には 39 件であったが、2016 年には 5 件に減少している。現在、DCI は世界有数のデータの検索手段であるものの、このレコード数だけで日本のデータ公開状況を判断することは難しいと言えよう。

表 1-1 Data Citation Index による分野別データレコード数

国／地域	2015		2016	
	All	Japan	All	Japan
遺伝学	2,277,115	42,196	2,965,110	61,183
生化学・分子生物学	1,781,274	42,390	2,001,247	51,368
その他の科学技術	503,387	0	882,401	21
結晶学	717,576	0	881,870	0
地質学	456,549	6	466,156	11
地理学	432,069	0	464,132	0
化学	42,207	2	144,316	3
海洋学	110,846	39	133,437	5
環境科学・生態学	32,674	0	85,751	17
分光学	79,303	19,485	79,877	19,652
調査対象レコード数	4,949,318	62,032	6,401,258	81,240
総レコード数	5,078,068	62,703	6,401,258	81,240
調査対象年	1995-2015		1996-2016	
調査日	2015/9/10		2016/9/4	

出典：Data Citation Index を用いて著者作成

以上のことから本研究は、質問紙調査によってデータ公開の実践状況を明らかにすることとした。調査にあたっては「内閣府報告書」で指摘されているように、“各分野の専門家・研究者、技術者の意見”（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015, p. 14）を広く収集して、データ公開に対する認識を明らかにすることを目指した。

¹⁵⁾ DCI は随時リポジトリを追加しているためレコード数が減ることはないと考えられるが、レコードの整理や分類の変更などによって、一部の分野や国・地域のレコード数が減少した可能性がある。

1.3.2 学術雑誌のデータ公開ポリシー調査

日本の研究者によるデータ公開の状況を、学術雑誌（以下、「雑誌」）によるデータ公開の要求状況と分野別に比較するため、データ公開ポリシー（以下、「ポリシー」）を調査した。データ公開の外的要因として雑誌のポリシーを対象とした主な理由は、研究者にとって論文の出版は主要な研究活動の1つであるため（Campbell, Pentz, & Borthwick, 2012）、影響が大きいと考えられることである。また、調査時点において日本の助成機関や研究機関はほとんどポリシーを策定していなかったため、分野別にデータ公開の要求状況を推測する指標として、雑誌のポリシーが最も適切であると判断した。

雑誌を通じたデータ公開は、データ公開を推進する組織や、一部の研究者によって重要であると認識されている。たとえば、論文にリポジトリで公開したデータの識別子を記載することや、論文の補足資料としてデータを掲載することの重要性が指摘されており（Whitlock, 2011; Anagnostou et al., 2015）、Royal Society (2012)、RECODE¹⁶⁾ (2014)、RDA Europe (2014)、Knowledge Exchange¹⁷⁾ (van den Eynden & Bishop, 2014) などの報告書で出版者¹⁸⁾にポリシーの策定を求める提言が行われている。生命科学分野の研究者を対象とした調査では、研究の推進に最も役立つとされたデータ共有ツールは「雑誌の補足データ」(58%)であり、「サードパーティのリポジトリ」(33%)を上回った（Pham-Kanter, Zinner, & Campbell, 2014）。神経画像分野（Breeze, Poline, & Kennedy, 2012）や材料科学分野（Gault, 2015）などの分野において、雑誌がデータ公開に重要な役割を果たしているとする論考もみられる。従って、データ公開の状況を捉えるために雑誌のポリシー（データ公開の要求状況）を調査することには一定の有効性があると考えた。

先行研究では、雑誌の要求が強いほどデータの公開率が高いという結果と、必ずしもそうではないとする結果が示されていた。Vines et al. (2013b) は雑誌のポリシーの強度を3段階（「義務化」、「推奨」、「なし」）に分類した上で12誌を調査した。その結果、要求が強い雑誌ほどデータの公開率が高かったため、ポリシーがデータ公開を強力に推進していると論じた。一方で、Spicer and Steinbeck (2018) によるメタボロミクス分野の調査では、雑誌のポリシーがあるにもかかわらず、データ公開が進んでいない場合もあることが示されていた。これらの結果をふまえて本研究では、分野別に雑誌のポリシーと実際の公開状況を比較して、両者に関連があるかどうかを明らかにすることとした。

¹⁶⁾ 欧州連合 (European Union, EU) による Horizon 2020 の助成を受けたプロジェクトの1つ。RECODE. <https://www.recodeh2020.eu/consortium>, (accessed 2018-10-31).

¹⁷⁾ 高等教育や研究のためのインフラ開発を目的とした英国の Jisc やドイツの DFG など6か国、6機関による組織。Knowledge Exchange. <http://www.knowledge-exchange.info>, (accessed 2018-10-31).

¹⁸⁾ 本研究が対象とする学術雑誌は、商業出版社だけではなく学協会などの団体も出版していることから、本論文では『図書館情報学用語辞典』（第4版）に準じて「出版社」ではなく、「出版者」と記す（日本図書館情報学会, 2013）。

1.3.3 比較分析の枠組み

ここで、質問紙調査および学術雑誌のデータ公開ポリシー調査による比較分析の枠組みについて述べる。まず、[RQ1]日本の研究者はどの程度データを公開しているのか、[RQ2]日本の研究者によるデータ公開の内的要因は何かを質問紙調査で明らかにする。その上で、日本の研究者の現状と特徴をより詳細に明らかにするために、まず、日本の研究者同士の比較を行う。具体的には、(1)分野間の比較、(2)論文のOAとの比較、(3)所属や年齢などの属性による比較、(4)データの公開経験がある回答者群とデータ公開経験がない回答者（非公開者）群の比較を行う。次いで、[RQ3]データ公開の外的要因はどのような状況にあるのか、この外的要因と日本の研究者によるデータ公開状況に乖離はあるのかどうかを明らかにするために、データ公開の主要な外的要因である(5)学術雑誌のポリシー（データ公開の要求状況）を調査して、日本の研究者による分野別のデータ公開率と比較する（図 1-2）。

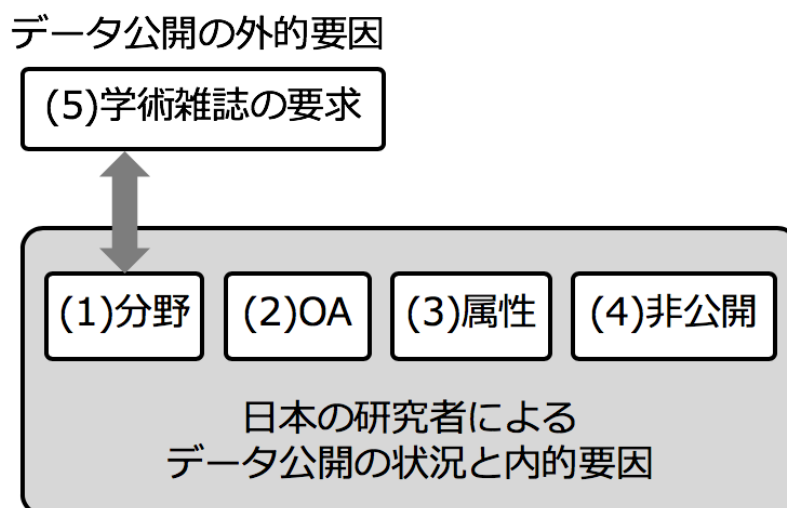


図 1-2 比較分析の枠組み

その結果、雑誌による要求よりもデータを公開している分野や、雑誌による要求に対してデータ公開が充分ではないと考えられる分野があれば、それぞれの認識や特徴を分析する。また、先行研究を参照して、国外の研究者によるデータ公開の状況や認識との比較を行う。以上の分析を通じて、日本の研究者によるデータ公開状況およびデータ公開に対する認識を相対的に明らかにすることを目指す。

1.4 研究の意義

1.4.1 学術情報研究への貢献

学術情報流通や学術コミュニティの研究領域において、論文や図書、特許などの学術情報の生産や公開状況、その国際比較といった研究は多数蓄積されてきた (Borgman & Furner, 2005; Steele, 2013; Weingart & Taubert, 2017)。データは学術情報の重要な一角を占めると論じられ (Borgman, 2008), “学術の一級市民 (first-class citizens of scholarship)” (Martone, 2014) と捉える論考もみられる一方で、出版物と比較するとまだ十分な調査は行われていない。

また、学術出版物を対象として、書誌データベースや引用索引データベースなどを用いて、分野による特性や差異、引用関係、共著関係などを明らかにするマクロな研究が盛んに行われてきた (Beaver & Rosen, 1978; Midorikawa, Ogawa, Saito, Kaneko, & Itsumura, 1984; Bornmann & Daniel, 2008; Greenberg, 2009; Wagner, et al., 2011)。しかし、データを対象とした研究については、特定の分野やプロジェクトを対象とした事例研究は多数行われているものの、分野横断での調査は充分に行われているとは言い難い (Edwards, Mayernik, Batcheller, Bowker, & Borgman, 2011; Weller & Monroe-Gulick, 2014; Silvello, 2017)。1 人の研究者が多数の事例研究を行った例として、Borgman (2015) による、自然科学、社会科学、天文学それぞれのビッグデータとスモールデータを扱うプロジェクトを対象とした調査がある。合計 6 領域¹⁹⁾には多彩なプロジェクトや実証実験が含まれており、Borgman はその半数に共同研究者として直接参加している。調査結果は構造化され、「メタデータ」や「知的財産権」など、主要なトピックを 6 領域にわたって比較することも可能である。それでもなお、結果は特定の事例から導き出されたものであるため、これを一般化することはできないと著者自身が指摘している。

以上のことから、データ公開の分野別の状況を捉えようとするところ、先行研究では未調査の分野を対象とすること、論文の OA とデータ公開の進展状況や課題の差異を明らかにすることが本研究の学術的貢献であると考えられる。

1.4.2 学術政策決定者や学術出版者への示唆

本研究の社会的な意義として、学術政策決定者や学術出版者の議論に資することを挙げる。本研究によって日本の研究者によるデータ公開の実践状況や認識、および分野による特性や差異が明らかになれば、適切かつ効率的な学術情報政策を立案するための一助となるのでは

¹⁹⁾ 具体的には、「天文学」(ビッグデータ)と「センサネットワークの科学技術」(リトルデータ、以下同様)、「インターネットサーベイとソーシャルメディア研究」と「社会技術研究」、「古典芸術と考古学」と「仏教研究」である。

ないだろうか。たとえば、多くの政策文書でデータ公開のためのプラットフォームの構築について触れられているが（内閣府, 2016; 文部科学省科学技術・学術審議会学術分科会第8期学術情報委員会, 2016; 日本学術会議, 2016; 内閣府, 2017), 日本の研究者が扱っているデータのプロファイルや機密情報などの問題については十分に明らかにされているとは言い難い。研究者がどのようなデータを扱い、どのような問題を抱えているのかを明らかにすることができれば、データ公開のためのプラットフォームの構築に際しての検討材料になると考えられる。また、多岐にわたる雑誌のポリシーを調査する過程で得られた知見は、学術出版者や助成機関、学術機関などが新たなポリシーを策定する際の基礎資料として有益であると考えられる。

分野の特性について、「内閣府報告書」は“研究分野によって研究データの保存と共有の作法に違いがあることを認識し、特性に応じたルール作りが必要である”（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015, p. 20）と指摘している。第5期科学技術基本計画では、“研究分野によって研究データの保存と共有の方法に違いがあることを認識する”（内閣府, 2016, p. 33）よう注意喚起している。しかし前項でも述べたように、分野を横断してその課題を明らかにするような研究は、現状では充分に行われていない。また、現時点では分野別のデータ公開状況についてデータベースなどを用いて把握するのは困難である。本研究で雑誌のポリシーを調査することによって、分野別のデータ公開要求の状況を把握し、日本の研究者のデータ公開状況と比較することができれば、優先的に支援が必要な分野の特定などに繋がるのではないかと考えられる。

1.4.3 研究データ管理での活用

国外の大学・研究図書館等は、助成機関によるデータ公開の義務化の動向を受けて、研究者のデータ公開を支援するための研究データ管理（Research Data Management, RDM）サービスを開始している（池内, 2014a; 池内, 2014b）。国内外のオープンサイエンス政策においても、データ公開のための人的支援やインフラの整備、人材育成が重要であると指摘されている。たとえばG7科学大臣会合の共同声明では、“オープンサイエンスが有効に活用され、全ての人がメリットを享受できるようにするために、国際的な協調や連携を推進して、デジタルネットワークの整備、人材の確保など、適切な技術やインフラを整備する”（G7 Science and Technology Ministers, 2016, p. 9）と述べられている。「内閣府報告書」に示された「政策立案及び実施における相関図」では、図書館・機関リポジトリ、ならびにデータセンター、国立情報学研究所（NII）が研究成果等の収集、OAの推進、共有されるデータの保存・管理を行う基盤として位置づけられた（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015）。「学術情報のオープン化の推進について（審議まとめ）」では、大学等に期待される取

1 章 研究背景と目的

り組みとして、“技術職員，URA²⁰⁾及び大学図書館職員等を中心としたデータ管理体制を構築し，研究者への支援に資するとともに，必要に応じて複数の大学等が共同して，データキュレーター等を育成するシステムを検討し，推進する”（文部科学省科学技術・学術審議会学術分科会第8期学術情報委員会, 2016, p. 13）と述べられている。「内閣府報告書」では，「5. 今後の検討課題等」として，「(6)データ駆動型の研究をサポートするサービスを企画，開発，運用する人材の確保」を挙げ，具体的に“科学技術分野ごとに異なるデータの属性，管理手法，利用者（特定のスキルの有無），利用局面等を理解できる人材の確保（データサイエンティスト，データキュレーター）”（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015, p. 22）を掲げている。

日本において研究データ管理体制を整備するためには，日本の研究者の実態を把握した上で，優先順位などを検討する必要があると考えられる。日本の研究者はどの程度データを公開しているのか，データ公開が要求されているのに対応できていない分野はあるのか，データ公開の具体的な障壁は何か，どのような支援が可能なのか，といった点を明らかにすることによって，適切かつ効率的な対応が可能になるのではないだろうか。本研究で明らかにしようとしている日本の研究者によるデータ公開の実践状況や認識に関する知見は，こうしたデータ公開支援のためのインフラ構築，人材育成，サービスの検討に資すると考えられる。

²⁰⁾ リサーチ・アドミニストレーター（University Research Administrator, URA）

1.5 用語の定義と分野分類

1.5.1 用語の定義

本研究で用いる用語の定義を示す。[1]～[4]の定義は、2016 年 11 月から 12 月にかけて実施した質問紙調査にも使用した。

[1] データ

研究のために収集・作成・観測したデジタルデータを指す。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しない。また、ゲノムデータ、地理情報、ソフトウェアコード、インタビューの録音と書き起こしなど、内容も限定しない。

※2.1.1 に示すように、「データ」の定義は極めて多様である。本研究は、多分野の研究者を対象とした調査を行うため、多様なデータを対象としている米国国立科学審議会（National Science Board, NSB）と OECD 原則の定義をベースとして素案を作成した（National Science Board [NSB], 2005, p. 18; OECD, 2007, p. 13–14）。質問紙調査の予備調査の結果（3.1.1）を参考に表現を修正した上で、最終的に上記の定義を用いることとした。

[2] カレントデータ

論文などの成果を発表済みの、最近の主要な研究 1 件のために収集・作成・観測したデータを指す。

※予備調査の結果、研究ごとに扱うデータの種類や量が異なる回答者が存在することが確認されたため、対象を限定して回答できるように定義した（詳しくは 3.1.1 で述べる）。

[3] データの提供

E-mail や USB フラッシュメモリ、クラウドサービス（Dropbox や Google Drive）などを使って、共同研究者を除く他者に渡す（共有する）ことを指す。特定の人以外はアクセスできない状態として、「データ公開」と区別する。

[4] データ公開

データをウェブサイトやリポジトリ、論文の補足資料などに掲載して、インターネットでアクセスして利用できる状態を指す。利用料金や利用者登録が必要な場合も含める。また、このような状態で公開されているデータを「公開データ」と呼ぶ。

※オープンサイエンス政策では、必ずしも無条件でのオープン化を求めておらず、特に国家安全保障、プライバシー、商業的利益などを含むデータは公開しなくても良いとする場合が多い。また、営利用禁止や利用料金、利用者登録など、何らかの条件を付けて公開しているデータも多い。こうしたデータ公開のレベルは、おおむね図 1-3 のように整理される。本研究は、図 1-3 の「オープン」と「条件付きオープン」の両方を「データ公開」として扱うこととした。

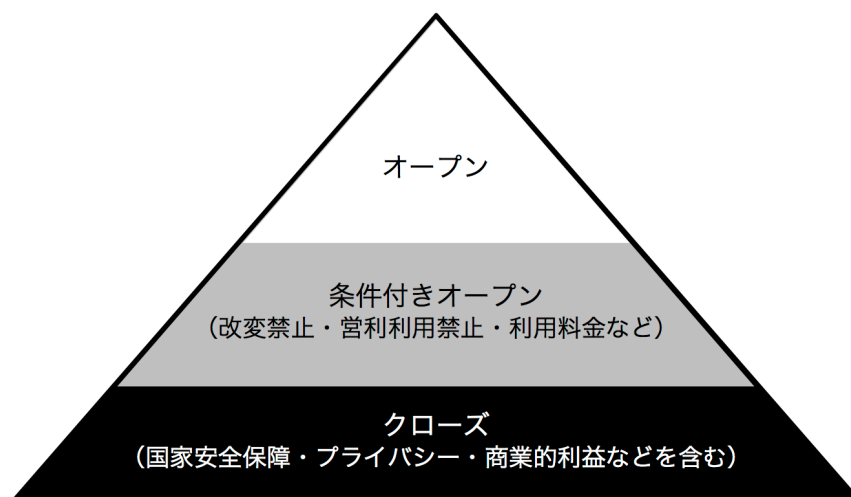


図 1-3 データの公開レベル

[5] 学術雑誌のデータ公開ポリシー

学術雑誌の投稿規定や執筆要綱など（以下、まとめて「投稿規定」と記す）に掲載された、データ公開に関する記述とする。データをリポジトリに公開してデジタルオブジェクト識別子（Digital Object Identifier, DOI）やアクセッション番号などの識別子を論文に示す方法と、論文の補足資料として電子版の雑誌にデータや識別子を掲載する方法（National Information Standards Organization [NISO] et al., 2013）を区別して扱う。本論文では、前者を「RP」（Repository Policy）、後者を「SP」（Supplemental Materials' Policy）と記す。

[6] オープンアクセス (OA)

論文がインターネットで公開され、読者は無料で読むことができる状態とする。たとえば、OA の雑誌で出版する場合や雑誌の OA オプションを選択した場合、雑誌等が一定期間経過後に論文を OA にする場合、著者が機関リポジトリやプレプリントサーバで論文を公開する場合を含む。いわゆるゴールド OA（雑誌等による OA 出版）とグリーン OA（著者によるセルフアーカイブ）の別は問わないこととした。

1.5.2 分野の分類

分野の分類は、Thomson Reuters 社²¹⁾の研究動向・統計データベースである Essential Science Indicators (ESI) の 22 分類²²⁾を用いた。ただし、質問紙調査の選択肢では回答者の負担を軽減するため、米国国立科学審議会 (National Science Board, NSB) の科学工学指標 (Science and Engineering Indicators, S&EI)²³⁾による 13 分類に「人文学」を追加した 14 分類を用いた。以下では、それぞれの選択理由と本研究における日本語名称について述べる。

[1] ESI の 22 分類

ESI は、過去 10 年分の Web of Science Core Collection (WoS) のデータに基づき、世界レベルでの分野ごとの研究動向、引用動向を提供するデータベースである²⁴⁾。ESI の 22 分類を採用した主な理由は、予備調査の結果、本調査にとって粒度が適切であると判断したこと、データ公開について分野横断で論じている主要文献 (Hey, Tansley, & Tolle, 2009; Royal Society, 2012; Kowalczyk & Shankar, 2013) で取り上げられている分類を含んでいること、計量書誌学の研究で分野間比較を行う際に用いられていること (たとえば, Bornmann and Daniel (2005) や Liu and Rousseau (2010) など)、および雑誌ごとに 1 種類の分類が付与されていること (小野寺, 2016) である。表 1-2 に、ESI の 22 分類と本研究で用いる日本語名称、および図表などのスペースが足りない場合に用いる略称を示す。

²¹⁾ 2016 年 10 月から Clarivate Analytics 社。2018 年 9 月現在、ESI の説明に変更はない。

²²⁾ “Field Definitions”. Clarivate Analytics. <http://archive.sciencewatch.com/about/met/fielddef/>, (accessed 2018-10-31).

²³⁾ “Appendix table 5-26: Fields and subfields of S&E publications data”. Science and Engineering Indicators 2012. National Science Board. “Appendix table 5-26: Fields and subfields of S&E publications data”. Science and Engineering Indicators 2012. National Science Board. <https://www.nsf.gov/statistics/seind12/append/c5/at05-26.pdf>, (accessed 2018-10-31).

²⁴⁾ “Essential Science Indicators”. Clarivate Analytics. <https://clarivate.com/products/essential-science-indicators/>, (accessed 2018-10-31).

1 章 研究背景と目的

表 1-2 ESI の 22 分類と本研究における日本語名称と略称

ESI の分野	日本語名称	略称
Agricultural Sciences	農業科学	農業科学
Biology & Biochemistry	生物学・生化学	生物学
Chemistry	化学	化学
Clinical Medicine	臨床医学	臨床医学
Computer Science	コンピュータサイエンス	CS
Economics & Business	経済学・経営学	経済学
Engineering	工学	工学
Environment/Ecology	環境・生態学	環境学
Geosciences	地球科学	地球科学
Immunology	免疫学	免疫学
Materials Science	材料科学	材料科学
Mathematics	数学	数学
Microbiology	微生物学	微生物学
Molecular Biology & Genetics	分子生物学・遺伝学	分子生物学
Multidisciplinary	複合領域	複合領域
Neuroscience & Behavior	神経科学・行動学	神経科学
Pharmacology & Toxicology	薬理学・毒性学	薬理学
Physics	物理学	物理学
Plant & Animal Science	植物・動物学	植物動物学
Psychiatry/Psychology	精神医学・心理学	精神医学
Social Sciences, General	社会科学	社会科学
Space Science	宇宙科学	宇宙科学

出典：“Essential Science Indicators”に基づき著者作成

[2] S&EI に基づく 14 分類

S&EI は、米国国立科学財団（National Science Foundation, NSF）の政策策定機関である NSB が、隔年で公表する科学・工学・技術分野の国際動向に関する報告書およびデータ集である（National Science Board [NSB], 2018）。1972 年版（1973 年刊行）から *Science Indicators* というタイトルで、1984 年からは *Science and Engineering Indicators* とタイトルを変更して継続的に出版されている（NSB, 2001, p. 20）。米国の政策立案のみならず、科学技術政策や分析に用

いられており (Adams & Griliches, 1996), 計量書誌学の研究で分野分類にも用いられている (たとえば, Archambault, Campbell, Gingras, and Larivière (2009) など)。質問紙調査の選択肢では回答者の負担を軽減するために, ESI の 22 分類を S&EI の 13 分類に「人文学」を追加した 14 分類に集約した。表 1-3 に S&EI に「人文学」を追加した本研究の分野分類と ESI との対応を示す。

表 1-3 S&EI の分類に基づく本研究の 14 分類と ESI の対応

S&EI の分野	本研究の 14 分類	ESI の分野
Agricultural Sciences	農業科学	農業科学 植物・動物学
Astronomy	天文学	宇宙科学
Biological Sciences	生物科学	生物学・生化学 免疫学 微生物学 分子生物学・遺伝学
Chemistry	化学	化学
Computer Sciences	コンピュータサイエンス (CS)	コンピュータサイエンス
Engineering	工学	工学 材料科学
Geosciences	地球科学	地球科学 環境・生態学
Mathematics	数学	数学
Medical Sciences	医学	臨床医学 神経科学・行動学 薬理学・毒性学
Physics	物理学	物理学
Psychology	心理学	精神医学・心理学
Social Sciences	社会科学	経済学・経営学 社会科学
Other Life Sciences	その他	複合領域
人文学【追加】		

出典：“Science and Engineering Indicators”と“Essential Science Indicators”に基づき著者作成

1.6 本論文の構成

図 1-4 に、本論文の構成を示す。1 章では研究背景としてデータ公開やオープンサイエンスの現状と課題について述べるとともに、研究目的と研究課題、研究方法、用語の定義など研究の枠組みを示した。2 章では、データ公開の動向や議論を概観した上で、本研究で実施する調査に関連する先行研究について述べる。3 章では、日本の研究者を対象とした質問紙調査によって[RQ1]日本の研究者はどの程度データを公開しているのかと、[RQ2]日本の研究者によるデータ公開の内的要因は何かを明らかにする。4 章では[RQ3]データ公開の外的要因はどのような状況にあるのかを明らかにするために、学術雑誌のデータ公開ポリシーを調査する。また、日本の研究者によるデータ公開状況との乖離はあるのかどうかを明らかにするために、3 章の結果と比較を行う。5 章では、3 章と 4 章の結果から導かれる結論として、日本の研究者によるデータ公開状況およびデータ公開に対する認識について述べる。また、本研究の新規性と限界、今後の課題を示す。

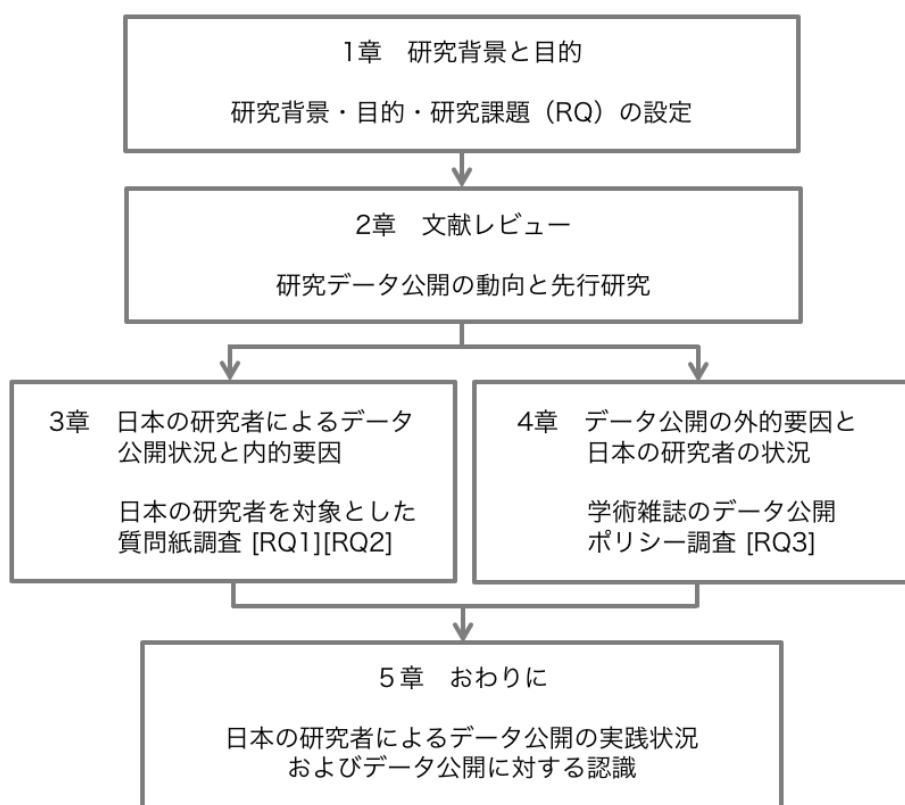


図 1-4 論文の構成

2 章 文献レビュー

本章では、本研究の対象であるデータそのものと、研究との関連や動向について述べる。続いてデータ公開の意義と期待される効果を整理した後、データ公開を推進する取り組みを概観する。その上で、先行研究としてデータ公開の実態調査と雑誌のデータ公開ポリシー調査に関する方法や知見について述べる。最後にまとめとして、先行研究で明らかにされた点と残された課題について述べる。

2.1 データと研究の関わり

まず、研究者が扱うデータとはどのようなものであるのか、さまざまな視点からの把握を試みる。そしてデータの共有や保存といった取り組みから、データを用いた研究の隆盛、データ公開とオープンサイエンスに至るまでの一連の動向を概観する。

2.1.1 データの多様性

本研究は分野横断の調査を行うために、できるだけ範囲を限定せずデータの定義を行った(1.5.1)。データの定義を検討する際に参考とした複数の文献を示す。また、データの分類やデータの規模とロングテールに関する記述を取り上げて、データが多様であることを示す。

[1] データの定義

研究データの定義やその対象とする範囲は文献によって異なる。ここでは、しばしば引用される定義として、米国国立科学審議会 (National Science Board, NSB)、経済協力開発機構 (Organization for Economic Co-operation and Development, OECD)、宇宙データシステム諮問委員会 (Consultative Committee for Space Data Systems, CCSDS)、そして日本学術会議の定義を挙げる。

NSB は、永続的なデジタルデータコレクションの報告書 (*Long-lived digital data collections: Enabling research and education in the 21st century*) において、コレクションに含まれるデータは多様であろうと指摘している。その上で、データには、数値、画像、ビデオ・オーディオストリーム、ソフトウェアとソフトウェアのバージョン情報、アルゴリズム、方程式、アニメーション、モデル／シミュレーションが含まれるとしている (National Science Board [NSB], 2005, p. 18)。

OECD の「公的資金による研究データへのアクセスに関する OECD 原則とガイドライン (*OECD principles and guidelines for access to research data from public funding*)」(以下、「OECD 原則」)では、“研究データ (research data)”は、科学研究の主要な情報源として利用される事実に基づく記録(数値のスコア、文字の記録、画像、音声)であり、研究結果の検証に必要であると科学コミュニティで一般に認められているものとされている。また、研究データセット (a research data set) は、調査の対象を体系的かつ部分的に表すものとされている。なお、ラボノート (laboratory notebooks)、予備的な分析、科学論文の草稿、将来の研究計画、査読、同僚との個人的なコミュニケーション、物理的な対象(例:実験室のサンプル、細菌株、マウスのような実験動物)は含まないとしている (Organization for Economic Co-operation and Development [OECD], 2007, p. 13–14)。

CCSDS は「OAIS 参照モデル (Reference Model for an Open Archival Information System, OAIS)」¹⁾において、データをコミュニケーション、解釈、処理に適した形式による再解釈が可能な情報の表現としている。具体例として、ビット列、数値の表、ページ上の文字、人の発話の音声記録、月の石の標本を挙げている (CCSDS, 2012, p. 1–10)。

日本学術会議は「回答: 科学研究における健全性の向上について」²⁾において、研究不正への対策のために保存すべき対象物を「情報やデータ (資料)」と「もの (実験資料や実験装置などの実体物)」に類型化している。さらに資料 (文書、数値データ、画像など) を、種類によって「デジタルデータ」(電子データ)と「アナログ資料」(紙媒体資料等)に分類している (日本学術会議, 2015, p. 5–6)。

このように、4 つの定義においてもデータの範囲は異なる。NSB の定義にはデータを利用するためのソフトウェアやそのバージョン情報が含まれるが他の 3 件には含まれておらず、CCSDS の定義にはデジタルデータのみならず物質 (標本) が含まれるが、他の 3 件には含まれていない。データの定義は広義のものから狭義のものまで多様であり、例示による定義には限界があるとも指摘されている (Borgman, 2015)。

[2] データの種類

デジタル保存 (digital preservation) に関する報告書では、研究データを(1)観測データ、(2)実験データ、(3)計算データ、(4)リファレンスデータセットの 4 種類に分類している (Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010, p. 56–57)。表 2-1 に、それぞれの解説と例をまとめた。

¹⁾ デジタル情報の長期保存アーカイブシステムの構築に関する国際標準規格 (ISO 14271:2003, ISO 14721:2012)。

²⁾ 2014 年に文部科学省科学技術・学術政策局長が公開した「研究活動における不正行為への対応等に関する審議について (依頼)」への回答である。

表 2-1 データの種類

種類	解説	例
観測データ Observational data	望遠鏡, 人口衛星, センサーネットワーク, 調査, その他の機器から収集される。歴史的記録や 1 回限りの現象であり, 多くの場合再現できない。	SDSS ³⁾ の天文データ ICPSR ⁴⁾ の社会調査データ
実験データ Experimental data	ハイスループットマシン (加速器など) から得られるデータ。データを再度収集することが実現不可能または非倫理的である。	臨床試験データ 創薬データ
計算データ Computational data	大規模なコンピュータシミュレーションによって生成される。データは再現可能。データマイニングなどに用いられる。短時間では再計算できない。	DOE ⁵⁾ や NSF ⁶⁾ が共有するスーパーコンピュータで実行
リファレンスデータセット Reference data set	高度にキュレートされたデータ。複数の科学コミュニティからの高い需要がある。	wwPDB ⁷⁾ の蛋白質構造データ PSID ⁸⁾ のパネル調査データ

出典：Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010) を参考に作成

また、データの種別を加工の度合いによって生データ (raw data) と加工データ (processed data) に分類する例もみられる。NASA の Earth Observing System Data and Information System (EOSDIS) によるデータは、加工の度合いによってレベル 0 から 1A, 1B, 2, 3, 4 までの 6 段階に分類されている⁹⁾。表 2-1 の例では、観測データは生データに近く、リファレンスデータセットは加工の度合いが高いといえる。分野や研究内容によって公開・保存すべきデータのレベルは異なるため、しばしば論点となると指摘されている (Doorn & Tjalsma, 2007)。

[3] データの規模とロングテール

図 2-1 は、データサイエンスの領域と特性、実現のためのレイヤーを示した欧州オープン

³⁾ Sloan Digital Sky Survey (SDSS). <http://skyserver.sdss.org/edr/en/>, (accessed 2018-10-31).

⁴⁾ Inter-university Consortium for Political and Social Research (ICPSR). <https://www.icpsr.umich.edu/icpsrweb/>, (accessed 2018-10-31).

⁵⁾ 米国エネルギー省 (Department of Energy, DOE). <https://www.energy.gov>, (accessed 2018-10-31).

⁶⁾ 米国国立科学財団 (National Science Foundation, NSF). <https://www.nsf.gov>, (accessed 2018-10-31).

⁷⁾ Worldwide Protein Data Bank (wwPDB). <https://www.wwpdb.org>, (accessed 2018-10-31).

⁸⁾ Panel Study of Income Dynamics (PSID). <https://psidonline.isr.umich.edu>, (accessed 2018-10-31).

⁹⁾ Data Processing Levels. Earth Science Data Systems (ESDS) Program. <https://earthdata.nasa.gov/earth-science-data-systems-program/policies/data-information-policy/data-levels>, (accessed 2018-10-31).

サイエンスクラウド（European Open Science Cloud, EOSC）¹⁰⁾プロジェクトによる図¹¹⁾である（池内, 2016）。この図には物理学、生命科学、地球科学などの研究領域が配置され、領域によって“科学活動の規模”が異なること、“科学のロングテール”にも差があることが示されている。つまり、分野によって扱うデータの量やデータの耐用年数（データの価値が認められる年数）が異なることが示唆されている。

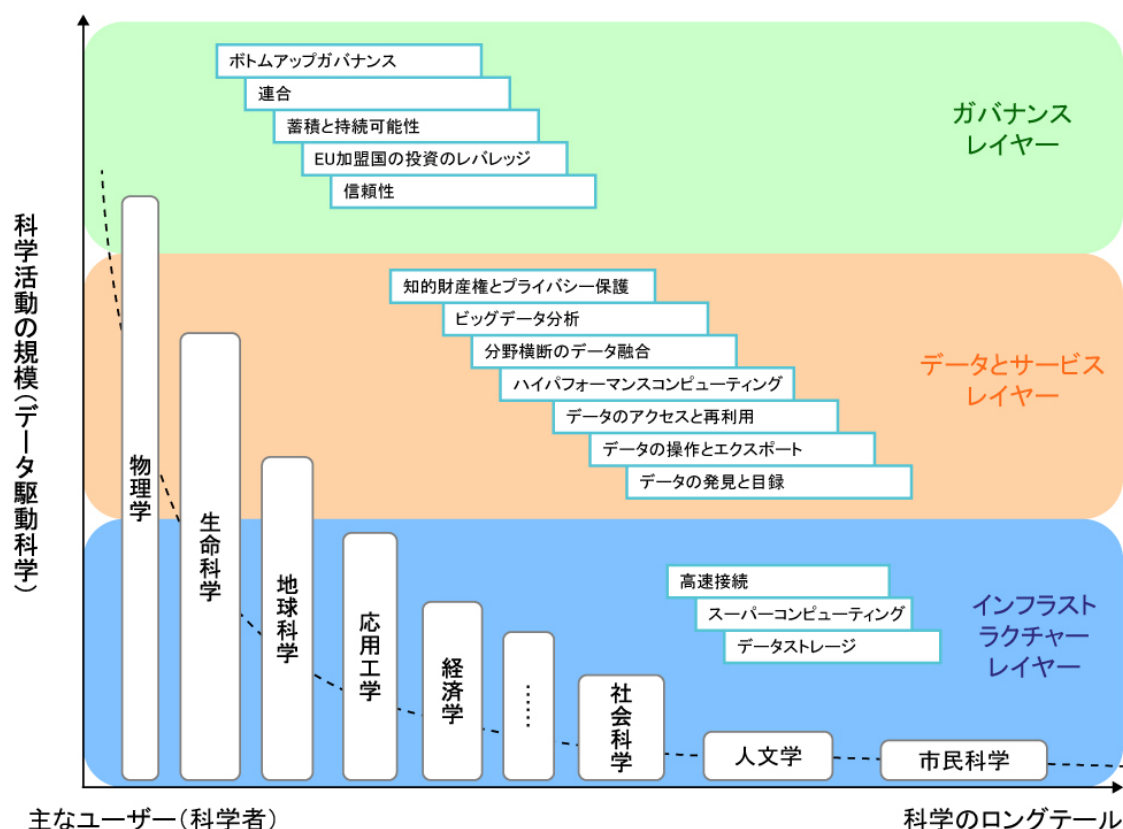


図 2-1 データサイエンスの領域と特性, 実現のためのレイヤー

出典：European Open Science Cloud. “Workshop on governance and funding for the EOSC”¹¹⁾

以上のようにデータの定義や種類は多様であり、分野や研究によって規模や扱いが異なる。データ公開の実現によって、こうしたデータを分野や国・地域を超えて再利用することや、異分野のデータを統合して新しい知見を生み出すことが期待されている。このことについては 2.2 で改めて述べる。

¹⁰⁾ EOSC は、欧州の全分野の研究者 170 万人と科学技術専門家 7,000 万人を対象としたデータ共有基盤として構想されている。

¹¹⁾ 2018 年 10 月現在、元の図は The European Open Science Cloud (Workshop on governance and funding for the EOSC) から入手できる。 <https://ec.europa.eu/research/openscience/pdf/eosc-workshop-06-2016/eosc.pdf>, (accessed 2018-10-31).

2.1.2 データ共有やデータ保存の取り組み

インターネット上で広くデータを公開する以前から、分野内でのデータ共有やデータ保存が行われてきた。本研究は、研究者のデータ共有などの経験とデータ公開には関連があるのではないかという前提に立ち、調査を行う。ここでは、データセンターやデータアーカイブ、デジタルキュレーションやデジタル保存の取り組みについて述べる。

[1] データセンター／データアーカイブ

InterPARES 2 Project (2008) の「科学データポータルによる保存の実践」に関する調査報告書によれば、データセンターやデータアーカイブといった組織は 1980 年から 2004 年の間に盛んに設立されてきた。その先駆けとして 1887 年に米国国立衛生研究所 (National Institutes of Health, NIH) が、1957 年に World Data Centre (WDC) が設立されている。その後、各国でさまざまな分野のコミュニティによってデータセンターやデータアーカイブが設立され、研究データの共有や保存の取り組みが進められてきた。国際組織として自然科学分野を対象とした国際科学会議 (International Council for Science, ICSU) による科学技術データ委員会 (Committee on Data for Science and Technology, CODATA) が、欧州の組織として社会科学データアーカイブ協議会 (Consortium of European Social Science Data Archives, CESSDA) があるなど、国際協力も行われている。

CODATA の設立にあたって、日本学術会議は内閣総理大臣宛に「自然科学におけるデータ確立集成活動の推進について」と題する勧告を行い、その重要性を説いた上でデータ活動推進のための方策を求めている (日本学術会議, 1968)。日本の動向について、前田 (2016) は 1960 年から 2015 年にかけて実施された研究データの整備に関する科学技術情報政策を調査した。その結果、日本では 1970 年代前半まではデータセンターに関する検討が行われており、1970 年代後半から 1990 年代までは数値データやファクト情報の枠組みによる検討が行われていたことを明らかにした。しかし、2000 年以降は研究データに関する包括的な方針や施策案が示されなくなると結論づけている。一方、2.3.2 に示すように 2015 年ごろからオープンサイエンス政策として、再び研究データに関する政策が打ち出されるようになってきた。

[2] データキュレーション／デジタル保存

データキュレーション (data curation) やデジタル保存などの取り組みでは、研究の成果物として生み出されるデータのみならず、研究やデータのライフサイクル全体を理解し、各段階での適切な処理や長期保存の実践について検討している。図 2-2 に Digital Curation Centre

(DCC) によるキュレーションのライフサイクルモデル¹²⁾を示す。この概念図にはデータの生成や取得、評価や選択と廃棄、取り込み、保存、蓄積、アクセス・利用・再利用、変換、そしてデータの生成や取得に至る一連の段階と、保存やキュレートといった実践のサイクルが示されている。

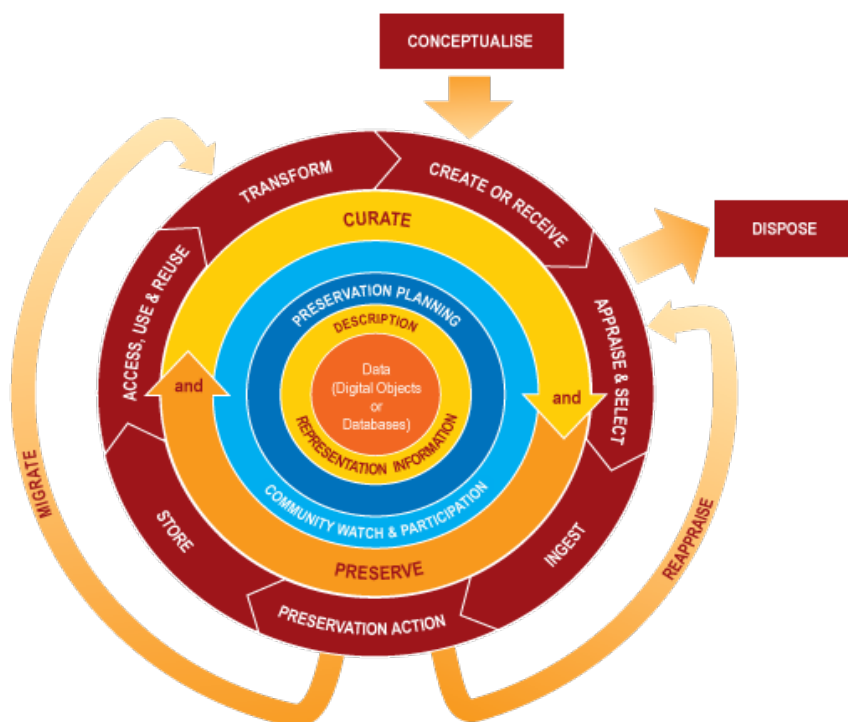


図 2-2 DCC Curation Lifecycle Model

各段階での実践に関する問題、たとえばリポジトリ、データ形式、メタデータ標準、マイグレーション、ライセンス、費用といった課題について、分野や国を超えた取り組みが行われている。こうしたデータ公開を推進するための活動については、2.3 で述べる。

2.1.3 データの増大とデータサイエンスの隆盛

情報通信技術の発展、特に計算能力の高速化、ネットワークの広帯域化、記憶装置の大容量化などによってデジタルデータの量が飛躍的に増大し、情報爆発 (data deluge) あるいはビッグデータの時代が到来した (Hey & Trefethen, 2003)。学術界では、大量かつ多様なデータを収集、蓄積、処理、分析することによるデータサイエンスが盛んになり、ゲノムデータや地

¹²⁾ DCC Curation Lifecycle Model. Digital Curation Centre (DCC). <http://www.dcc.ac.uk/resources/curation-lifecycle-model>, (accessed 2018-10-31).

球観測データなどのビッグデータを統合し、異分野で活用することによって新たな成果を生み出すことが期待されている。さらに、これらのデータをビジネスやイノベーションの創出につなげることによって経済効果を生むことも期待されている。Microsoft Research の Jim Gray は、こうした科学の変化を第4パラダイム¹³⁾と名付けた (Hey et al., 2009)。

2000 年ごろから各国の学術情報政策により、e-Science やデータ集約型研究 (data intensive science) が推進されてきた。米国はサイバーインフラストラクチャ計画、英国は e-Science 政策、豪州は全国データシステム、欧州委員会は FP7 による e-infrastructure の整備が進められ、国や地域レベルでデータサイエンスの基盤が整えられてきた。日本においてもビッグデータやデータサイエンスに対応するためのアカデミッククラウドの整備に関する議論が重ねられ、2012 年には提言が公開された (文部科学省アカデミッククラウドに関する検討会, 2012)。

2.1.4 データ公開からオープンサイエンスへの展開

研究データの共有や公開、データ量の増加、新たな分析手法の開発、それに伴う研究方法の変化や発展 (OECD, 2014)、市民科学による学術コミュニティの拡がりといった一連の潮流は、Science 2.0、オープンスカラシップ、オープンリサーチ、参加型サイエンス、e-Science など、さまざまな名称で呼ばれてきた。しかし、2014 年 7 月から 9 月にかけて欧州委員会 (European Commission, EC) が実施したパブリックコメントによって、研究者をはじめとするステークホルダーが最も好んでいる名称は、オープンサイエンスであることがわかった (European Commission [EC], 2015)¹⁴⁾。この頃からオープンサイエンスという名称は国内外の政策や報告書で用いられるようになり、徐々に定着してきたといえる。

オープンサイエンスという用語は、公的な研究によって生み出された科学的知識は公共財 (public goods) であるという考えに基づき、経済学者の Paul David が案出したとされている (OECD, 2015)。「OECD 原則」では、研究データとは別に「公的資金による研究データ」を定義している。具体的には、政府機関や部門によって行われた研究から得られたデータ、あるいは政府によって提供された公的資金を利用した研究から得られたデータとしている (OECD, 2007, p. 14)。David (2003) は、こうしたデータを誰でも追加費用なく利用できるようにすることによって、より大きな社会的利益を生み出すという考えを示している。

¹³⁾ 第1パラダイムは数千年前の経験科学 (empirical science) を指し、自然現象を観測し、記述する段階である。第2パラダイムは数百年前の理論科学 (theoretical science) を指し、ケプラーの法則、ニュートン力学、マクスウェルの方程式などを用いる段階である。第3パラダイムは数十年前の計算科学 (computational science) を指し、より複雑な現象をシミュレーションによって予測する段階である。そして第4パラダイムは理論、実験、シミュレーションを統合した e-Science あるいはデータ中心の科学 (data centric science) の段階である。計測器によってキャプチャされたデータやシミュレータによって生成されたデータがソフトウェアによって加工され、コンピュータに情報や知識が保存される。科学者はデータを管理し、統計を用いてデータベースやファイルを分析する。

¹⁴⁾ Science 2.0 という名称のパブリックコメントであったにもかかわらず、回答者の選択率が最も高かったのは「オープンサイエンス」(43%) であった。

武田（2018）はオープンサイエンスの歴史的経緯を俯瞰した上で、そのルートとして(1)科学そのもののオープン化, (2)大学・研究図書館を通じた OA 運動, (3)政府のオープンデータ, (4)市民科学があると論じている。船守（2017）は、オープンサイエンスを大学等の学術機関による対応可能性を検討する観点から、「研究成果や研究資料の公開・共有」、「新たな学術の次元の追求」、「社会との協働」に分類している。以下では、「オープン」と「オープンサイエンス」の定義について述べる。

[1] オープンの定義

データの定義と同様に、「オープン」の定義も多様である。2005 年に Open Knowledge Foundation (OKF)¹⁵⁾は「オープン」の定義を試みた。その最新版である Open Definition 2.1 の要約では、オープンとは、誰もがあらゆる目的で自由にアクセス、利用、改変、共有できることを意味するとしている (Open Knowledge International, n.d.)。

「OECD 原則」による公開 (openness) の定義は、可能な限り低いコスト、できれば流通の限界費用で、国際的な研究コミュニティが同じ条件 (equal terms) でアクセスできることである。なお、公的資金による研究データへの OA は、簡単で、タイムリーで、ユーザーフレンドリーであり、できればインターネットベースであるべきとされている (OECD, 2007, p. 15)。

OKF や「OECD 原則」の定義は理想的な状態を表しているが、研究データにはライセンスや機密、プライバシー情報など法的、倫理的、経済的な問題が含まれるため、現状ではデータによって制限がかけられる場合がある。また、公開の範囲もいくつかの段階に分けられる。Whyte and Pryor (2011) は、文献調査とインタビュー調査に基づき、公開 (open) の程度を 6 つに分類した。すなわち、(1)プライベート管理 (研究グループ内)、(2)共同的な共有 (プロジェクトなどのコンソーシアム内)、(3)同僚 (peer) との交換 (研究者ネットワーク内)、(4)透明なガバナンスのための共有 (所属機関や助成機関の要求に応じた外部公開)、(5)コミュニティにおける共有 (研究コミュニティ内)、(6)公的共有 (制限がほとんどない状態で誰でもアクセス可能) である。

[2] オープンサイエンスの定義とステークホルダー

表 2-2 に主なオープンサイエンスの定義を挙げる。その指し示す範囲を正確に捉えることは難しいが、表 2-2 では、(1)オープンアクセス (Open Access, OA) とデータ公開を対象とする定義と、(2)OA とデータ公開以外にもオープンピアレビューなど多様な研究活動を例示している定義にわけて示した。

¹⁵⁾ 2004 年に英国で政府データの公開や利用の促進を目指して設立された。現在の名称は Open Knowledge International。 <https://okfn.org>, (accessed 2018-10-31).

表 2-2 オープンサイエンスの定義

出典	定義
内閣府 (2016)	“オープンアクセスと研究データのオープン化（オープンデータ）を含む概念である”
G7 Science and Technology Ministers (2016)	“幅広い分野の公的資金による研究成果（論文や関連するデータセット等）に学術関係者だけでなく，民間企業や一般市民が，広く利用・アクセスできるようにするものである”
EC (2018)	出版物の OA，研究データ公開，オープンな学術コミュニケーション（オープンピアレビュー，雑誌のオープンピアレビューポリシー，オルトメトリクスなど）
OECD (2015)	“論文や研究データといった公的資金による研究から得られた結果を，今後の研究促進のために，無制限または最小限の制限のもとにデジタル形式でオープンにアクセスできることを目指す，研究者，行政，研究助成機関，あるいは科学界自体による取り組み” “オープンピアレビュー，オープンリサーチのノート，研究素材へのオープンアクセス，オープンソフトウェア，市民科学，研究資金のクラウドファンディングもオープンサイエンスシステムの構造の一部である”
Vicente-Saez & Martinez-Fuentes (2018)	共同のネットワークを通じて共有され，開発される，透明でアクセス可能な知識 OA，オープンデータ，オープンコード，オープンピアレビュー，市民科学，オープンラボノートなど

(1)OA とデータ公開を対象とする定義として，第 5 期科学技術基本計画は，オープンサイエンスを“オープンアクセスと研究データのオープン化（オープンデータ）を含む概念である”（内閣府，2016，p.32）としている。2016 年の G7 科学技術大臣会合による共同声明もこれに近く，“幅広い分野の公的資金による研究成果（論文や関連するデータセット等）に学術関係者だけでなく，民間企業や一般市民が，広く利用・アクセスできるようにするものである”（G7 Science and Technology Ministers, 2016, p.9）としている。

(2)OA とデータ公開以外も含めている例として，まず，EC によるオープンサイエンスの成果を追跡するための枠組みであるオープンサイエンスモニター（Open Science Monitor）¹⁶⁾を挙げる。オープンサイエンスモニターは，出版物の OA（プレプリント，助成機関や雑誌の OA ポリシー）と研究データ公開に加えて，オープンな学術コミュニケーションもその範囲に含

¹⁶⁾ https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en, (accessed 2018-10-31).

めている。オープンな学術コミュニケーションの要素として、オープンピアレビュー、雑誌のオープンピアレビューポリシー、オルトメトリクス (altmetrics) などを挙げている。

OECD (2015) の報告書 *Making Open Science a Reality* では、オープンサイエンスの統一的な定義はまだ確立していないと断りつつも、“論文や研究データといった公的資金による研究から得られた結果を、今後の研究促進のために、無制限または最小限の制限のもとにデジタル形式でオープンにアクセスできることを目指す、研究者、行政、研究助成機関、あるいは科学界自体による取り組み”であり、“オープンピアレビュー、オープンリサーチのノート、研究素材へのオープンアクセス、オープンソフトウェア、市民科学、研究資金のクラウドファンディングもオープンサイエンスシステムの構造の一部である”としている。オープンサイエンスモニターに近い内容であるといえよう。

Vicente-Saez and Martinez-Fuentes (2018) は、現時点での厳密かつ統合的なオープンサイエンスの定義を行うために、2006 年から 2016 年に刊行されたオープンサイエンスに関連する文献を Web of Science, Scopus, 国際データベース¹⁷⁾から抽出して、システマティックレビューを行った。その結果、オープンサイエンスとは、共同のネットワークを通じて共有され、開発される、透明でアクセス可能な知識であると結論づけている。その要素として、オープンサイエンスモニターや OECD (2015) と同様にオープンピアレビューや市民科学を挙げているほか、さらにオープンコードやオープンラボノートなども含めている。

広義のオープンサイエンスは、研究活動とその成果に、誰もが自由にアクセスして使えることと捉えられる。研究成果にはデータ、コード、ラボノートなどや出版物 (学術論文、書籍、プレプリントなど) が含まれ、研究活動には査読、評価、資金獲得などが含まれる。ステークホルダーには、研究者だけではなく、市民、政府、企業などが含まれ、その活用は学術に限らず、商用利用など自由な利用が想定されている。狭義のオープンサイエンスの定義は OA とデータ公開であると捉えられるが、いずれの場合も、OA とデータ公開を含む概念として規定されている。

OECD の報告書では、オープンサイエンスの主要な当事者として、(1)研究者、(2)政府省庁、(3)研究資金提供機関、(4)大学や公的研究機関、(5)図書館、リポジトリ、データセンター、(6)民間の非営利団体、財団、(7)民間の学術論文出版者、(7)企業、(8)超国家機関 (OECD や UNESCO 等) を挙げている (OECD, 2015)。北本 (2015) は、オープンサイエンスに関わる多様なステークホルダーと、それぞれの目的や誘因のあり方を「同床異夢」と表現している。

2.2 データ公開の意義と効果

多様なステークホルダーが関わるオープンサイエンスの文脈では、データ公開によって学

¹⁷⁾ 例として、欧州連合、国際連盟、OECD、世界銀行が挙げられている。

術界のみならず社会にもさまざまな効果をもたらすことが期待されている (OECD, 2015)。ここでは 1 章でも言及した, [1]研究の効率化, [2]異分野データの活用と共同研究の推進, [3]研究の透明性の向上, [4]データの長期保存, [5]市民への研究成果の還元, [6]イノベーションの創出と経済効果, [7]市民科学の進展について述べる。

[1] 研究の効率化

2003 年にデータ共有ポリシーを策定した米国国立衛生研究所 (National Institutes of Health, NIH) は, データ共有の利点として研究やデータの重複を防ぐことを挙げている (NIH, 2014)。公開されたデータを再利用することによって, 後続研究を効率的に進めることができる。また, 失敗した研究のネガティブデータを公開することによって, 研究の無駄を省く試みも行われている (Nosek, Spies, & Motyl, 2012)。

2012 年の OECD 主要科学技術指標によれば, 各国の研究開発に対する国内総支出 (購買力平価レート) は, 米国が 4,535 億ドル (対 GDP 比 2.8%), 日本は 1,517 億ドル (同 3.4%), 英国は 391 億ドル (同 1.7%), そして欧州連合 (European Union, EU) 15 カ国は 3,175 億ドル (同 2.1%) であった (OECD, 2012)。研究プロセスの重複を防ぐことは, 研究開発費の二重投資を減らすことになり, 財源である税金の節減にも繋がる。Houghton and Gruen (2014) は, オーストラリア国立データサービス (Australian National Data Service, ANDS) への報告書で, 同国の公的資金による研究データの価値を年間 19 億ドルから 60 億ドルまで, リポジトリの価値を年間 5.5 億ドルから 18 億ドルまでと見積もっている。

[2] 異分野データの活用と共同研究の推進

分野を超えたデータの利活用や, 異分野データの統合による新たな成果の創出が期待されている (Edwards et al., 2011)。G8 科学大臣共同声明では, “社会科学, 人文科学, 自然科学, 生命科学, 環境科学にわたる学際的で分野横断的なアプローチを取ることの重要性”が強調されている (G8 Science Ministers, 2013)。分野や国境を超えたデータの利活用に向けて, 研究データ同盟 (Research Data Alliance, RDA) (Treloar, 2014) や DataCite¹⁸⁾ (福山, 2015) といった国際組織が設立され, 既存のデータセンターや国立・研究図書館, 出版社, 研究者などと協力しながら基盤整備を進めている。日本における実例として, 2006 年に設立された Data Integration and Analysis System (DIAS) は, 地球規模の観測データを社会経済情報などと融合して学際研究を推進している¹⁹⁾。

¹⁸⁾ 研究データへの DOI 発行やメタデータスキーマの提供によって, データの発見や引用, 論文へのリンクなどを容易にすることを目指している。“What do we do?”. DataCite. <https://www.datacite.org/about-datacite/what-do-we-do>, (accessed 2018-10-31).

¹⁹⁾ “DIAS とは”. DIAS: データ統合・解析システム. <http://www.diasjp.net/about/>, (accessed 2018-10-31).

Prosser (2013) は、近年、助成機関は学際研究と国や機関を超えた共同研究を奨励していると指摘している。実際に、複数の国や機関による共同研究は被引用数や評価が高いこと (Sanfilippo, Hewitt, & Mackey, 2018)、また、流動性の高い研究者は生産性も高いことなどが明らかにされている。こうした理由から、国際共著率は研究のベンチマークとして用いられ (Luukkonen, Tijssen, Persson, & Sivertsen, 1993; Sonnenwald, 2007; Abbasi, Altmann, & Hossain, 2011)、世界大学ランキングにおいても国際共同研究は評価の指標として用いられている。現状では、こうした共同研究の状況は国や分野によって異なり、限定的であると指摘されている (Hottenrott & Lawson, 2017)。しかしながら公開したデータを契機として、共同研究や国際共著などが推進されること (Procter et al. 2010; Borgman, Wallis, & Mayernik, 2012) が期待されている (“Datas shameful neglect”, 2009; Kowalczyk & Shankar, 2013)。

[3] 研究の透明性の向上

米国研究評議会 (National Research Council, NRC) は、2002 年の報告書で、ある研究が再現可能であることはゴールドスタンダードであり、それによって科学的主張の信頼性が判断されるとしている (NRC & Global Affairs, 2002)。研究不正の防止 (Ioannidis, 2014)、特に人類の健康や生命に直接関わる医学分野におけるデータの改竄や捏造の防止 (Doshi, Goodman, & Ioannidis, 2013; European Medicines Agency [EMA], 2014) は、学術コミュニティだけではなく社会全体にとっても重要な意味をもつ (Arzberger, 2004; Guttmacher, Nabel, & Collins, 2009)。Naudet et al. (2018) は、2013 年から 2016 年に *BMJ* と *PLOS Medicine* に掲載されたランダム化比較試験 (Randomized Controlled Trial, RCT) を含む論文が公開しているデータを用いて結果の再現を試みた。その結果、完全に再現できたのは 17 件中 14 件 (82%) であった。研究の再現性は、心理学 (Klein et al., 2014; Maxwell, Lau, & Howard, 2015) や経済学 (Kane, 1984; Dewald, Thursby, & Anderson, 1986; McCullough, McGeary, & Harrison, 2006) 分野において重要課題とされ、しばしば“再現性の危機 (replication crisis)”の問題として論じられてきた。近年、再現性の危機の問題は、多分野の研究者が認識していることも明らかにされている (Baker, 2016)。

Retraction Watch²⁰⁾や PubPeer²¹⁾といった学術コミュニティサイトにおいても、論文の根拠データの追試によって研究の誤りを指摘する事例がみられる。創設者である Ivan Oransky (Retraction Watch) や Brandon Stell (PubPeer) は、公開データの再分析や再現性の重要性を指摘している (Didier & Guaspare-Cartron, 2018)。

²⁰⁾ 学術雑誌から撤回された論文についての情報や議論を掲載しているブログサイト。Retraction Watch. <http://retractionwatch.com>, (accessed 2018-10-31).

²¹⁾ 研究者が特定の学術論文について公開で議論するウェブサイト。PubPeer. <https://pubpeer.com>, (accessed 2018-10-31).

[4] データの長期保存

データの多くは研究終了後に失われてしまうが、データ公開のために再利用可能な形でリポジトリに登録することによって、長期間保存されることが期待される (Karasti, Baker, & Halkola, 2006; Abrahamson, Bollen, Gutmann, King, & Pienta, 2009; Conway, Giaretta, Lambert, & Matthews, 2011)。2.1.2 で述べたようなデジタル保存の取り組みにおいても、学術出版物 (scholarly discourse)、商用の文化的コンテンツ、ウェブコンテンツとともに研究データが取り上げられ、永続的な保存のための提言が行われている (Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2010)。国際イニシアティブの FORCE11²²⁾は、データが長期にわたって再利用できるように、Findable (見つけられる)、Accessible (アクセスできる)、Interoperable (相互運用できる)、Re-usable (再利用できる) 状態で公開する指針として、FAIR 原則 (FAIR Data Principles) を提唱した (Wilkinson et al., 2016)。日本では、バイオサイエンスデータベースセンターによる解説が公開されている (大波ら, 2018)。

Vines et al. (2014) は 1991 年から 2011 年に出版された 516 報の論文の著者を対象として、E-mail によるデータの入手を試みた。論文年齢 (出版後の経過年数) が高くなるにつれて著者への連絡やデータの入手が難しくなり、22 年前の論文 22 報のうち、データを受け取れたのは 1 報 (4.5%) であった。調査の結果から、データの生存確率は 1 年に 17% ずつ減少すると結論づけられている。しかし、データを適切な形でリポジトリに登録すれば、長期にわたって保存され、再利用も可能になると考えられる。実際に、Vines et al. (2014) の調査データと分析に用いた R のコードは、データリポジトリの Dryad²³⁾で、永続的な識別子であるデジタルオブジェクト識別子 (Digital Object Identifier, DOI)²⁴⁾を付与して公開されている (Vines et al., 2013a) ため、長期にわたって保存されると考えられる。

[5] 市民への研究成果の還元

研究データを公共財として市民に公開することも期待されている (Baron, 1988; Rodwin & Abramson, 2012; André, 2016)。科学的な手法による公開データを用いて、地球規模の課題から国や地域レベルの課題まで、さまざまな問題を検討することによって、いわゆるエビデンスに基づく政策決定 (evidence-based policy making) を実現することが可能になると指摘されている (Sanderson, 2002; Choi et al., 2005; Howlett, 2009)。

市民による教育での活用なども期待されている (Salmi, 2014)。UNESCO の調査によれば、

²²⁾ 科学的知識の共有を推進する国際イニシアティブ。FORCE11 (Future of Research Communications and e-Scholarship). <https://www.force11.org>, (accessed 2018-10-31).

²³⁾ Dryad. <http://datadryad.org>, (accessed 2018-10-31).

²⁴⁾ DOI とは、有形・無形のオブジェクトに付与される永続的かつ一意の番号である。リンク切れを防ぎ、恒久的なアクセスを可能にすることを目的として開発された。

PubMed Central²⁵⁾で公開されている OA 論文の一日あたりのユニークユーザ 42 万人のうち、40%が市民、大学が 25%、企業が 17%であった (Swan, 2012)。データは論文とは異なりフォーマットが一定ではないが、徐々に活用事例も増えている。たとえば、13 歳から 18 歳を対象とした科学コンテストである Google Science Fair²⁶⁾では、参加者の多くが公開リポジトリのデータを用いた発明²⁷⁾を行っている (池内, 2017)。日本においても、2020 年から小学校でのプログラミングが必須化され (文部科学省, 2016)、2016 年から総務省統計局 (n.d.) によるオープンデータを活用するためのデータサイエンスや統計に関するオンライン講座が定期的に関講されるなど、公開データの活用を推進するための取り組みが行われている。

[6] イノベーションの創出と経済効果

研究データの商用利用による経済的な効果も期待されている。Caulfield, Harmon, & Shawn (2012) は、ゲノム研究に関するカナダ、米国、および英国のガイドラインを分析した上で、研究を商業化することへの圧力とオープンサイエンス政策は必ずしも折り合わないとはいえず、より包括的なイノベーションのフレームワークにおける補完的な要素と考えられると論じている。また、2017 年 6 月にデータジャーナルの *Scientific Data* (Springer Nature) は、同誌に投稿されたデータセットは商用利用の制限を認めないとする方針を公開した (機密データは除く) (Scientific Data, 2017)。間接的な背景として、産業界におけるビッグデータの収集とそれを分析するデータサイエンティストの活躍や (Davenport & Patil, 2012)、政府機関によるオープンデータの拡がり (Office of Science and Technology Policy [OSTP], 2014)、オープンデータを活用した新たなサービスの開発 (Lakomaa & Kallberg, 2013) などがある。つまり、社会全体でデータの利活用が盛んになっていることも、研究データの幅広い利用を後押ししていると考えられる。

[7] 市民科学

研究者ではない市民がデータの採取や整理、分析や議論に参加する市民科学が盛んに行われている (Curry, Freitas, & O'Riain, 2010; Heckerman, Takeda, Emmott, Hey, & Xu, 2012; Bisol et al., 2014)。数万人規模の市民がデータを介して研究に関わる事例もあり、クラウドサイエンスとも呼ばれる。研究者は新たな研究方法による発見や解析が可能になり、市民には科学リ

²⁵⁾ 米国生物工学情報センター (National Center for Biotechnology Information, NCBI) の生物医学・生命科学論文アーカイブ。現在の名称は PMC。https://www.ncbi.nlm.nih.gov/pmc/, (accessed 2018-10-31).

²⁶⁾ Google Science Fair https://www.google-sciencefair.com/, (accessed 2018-10-31).

²⁷⁾ 2012 年に Google Science Fair でグランプリを受賞した Brittany Wenger 氏は、公開データを用いて乳がんと混合系統白血病の診断方法を開発した。Wenger 氏は、自身で Cloud4Cancer Breast Cancer Detection (http://www.cloud4cancer.appspot.com) というサイトを立ち上げ、世界中の専門家からデータの提供を受けて診断の予測精度を高めている。

テラシーの向上や才能の発見といった利点がもたらされている (Bonney et al., 2009; Scientize, 2014; 林, 2015)。

2.3 データ公開を推進する取り組み

このように、データ公開による意義や効果が期待されることから、データ公開を推進する取り組みが行われてきた。ここでは、まず学術コミュニティの動向、次いでオープンサイエンス政策の動向について述べる。最後に、これらを契機とした図書館等による研究データ管理サービスを取り上げる。

2.3.1 学術コミュニティの動向

表 2-3 に、データ公開に関する学術コミュニティの主な動向を示す。表にはここまでで言及してきた RDA の設立や FAIR 原則 (FAIR Data Principles) なども含めた。

表 2-3 データ公開に関する学術コミュニティの主な動向

2009 年	DataCite 設立
2010 年	Panton Principles: Principles for Open Data in Science
2012 年	<i>Science as an Open Enterprise</i> (Royal Society) BMJ Open Data Campaign LERU statement on Open Research Data Data Citation Index (DCI) 公開 (Thomson Reuters 社)
2013 年	研究データ同盟 (Research Data Alliance, RDA) 設立
2014 年	Joint Declaration of Data Citation Principles (FORCE11) データジャーナル <i>Scientific Data</i> (NPG), <i>Data in Brief</i> (Elsevier 社) 創刊 Making Data Count (MDC) プロジェクト開始
2016 年	Amsterdam Call for Action on Open Science FAIR Data Principles (FORCE11) 研究データ利活用協議会 (RDUF) 設立
2018 年	Google Dataset Search (ベータ版) 公開

データ公開を牽引する報告書は多数公開されているが、ここでは代表的なものとして 2012 年に英国の Royal Society が公開した、*Science as an Open Enterprise* と題する報告書を取り上げる (Royal Society, 2012)。当該報告書は、科学の発展と社会的、経済的な利益の最大化のためには倫理や機密に触れない限り、あらゆる分野の研究データが公開され、活用可能にする

ことが望ましいとして、データ公開を後押ししている。同様の指摘は、科学データに関する専門家グループ（High Level Expert Group on Scientific Data）による EC への報告書 *Riding the wave: How Europe can gain from the rising tide of scientific data* にもみられる（European Union, 2010）。

データ公開に関する組織的な活動を挙げる。2012 年に OKF¹⁵⁾の協力を得て公開された Pantan Principle は、パブリックドメインでデータを公開すること、すなわち、データに関するさまざまな権利を放棄して利活用を促進することを強く提案している（Murray-Rust, Neylon, Pollock, & Wilbanks, 2012）。*BMJ (British Medical Journal)* は、2012 年からオープンデータキャンペーンを展開して、臨床試験データの公開を推奨している²⁸⁾。また、欧州研究大学連盟（League of European Research Universities, LERU）は、*LERU statement on Open Research Data* を公開して研究データ公開の推進を表明した（League of European Research Universities [LERU], 2012）。

データを他者が再利用できる形で公開することをデータ出版（data publishing）と呼び、論文の出版と同様に扱ったり、引用したり、データベースで検索可能にする動きもみられる。データ出版について、論文にかかわって公開したデータの識別子とデータに関する記述（data descriptor）を掲載するデータジャーナルの刊行や、学術雑誌へのデータペーパーの掲載が増加しており（Candela, Castelli, Manghi, & Tani, 2015）、研究成果をデータの形で広く共有することが増えている（Borgman, 2015）。2014 年にはデータジャーナルの *Scientific Data*（Nature Publishing Group, NPG）や *Data in Brief*（Elsevier）が相次いで創刊された。

論文を引用するように、データを研究に用いた際に典拠を示す、データ引用（data citation）を促す取り組みも進められている。その目的や効果は、(1)データの作成者にクレジットを与えること、(2)データの引用状況を測定することによって研究の影響度を明らかにして、公開のインセンティブとすること、(3)データへのアクセスを確実にすること、(4)追試や検証を可能にして研究の透明性を向上させることなどであり、CODATA と国際科学技術情報会議（International Council for Scientific and Technical Information, ICSTI）、RDA は、データ引用の標準化や普及に努めてきた（CODATA-ICSTI Task Group on Data Citation Standards and PractOut of Cite, Out of Mind: The Current Sices, 2013）。これらの機関や学術出版社とも連携して、FORCE11 は 2014 年にデータ引用原則の共同宣言（Joint Declaration of Data Citation Principles, JDDCP）を（FORCE11, 2014）、2016 年にはソフトウェアの引用指針（Software citation principles）を公開した（Smith, Katz, Niemeyer, & FORCE11 Software Citation Working Group, 2016）。JDDCP には各国・地域の学術機関や出版社が賛同しており（Staar et al., 2015）、2016 年 11 月には Elsevier 社が 1,800 を超える雑誌に採択した（Elsevier, 2016）。

こうしたデータやデータ引用の検索ツールの開発も進められてきた。2012 年 11 月に

²⁸⁾ “BMJ Open Data Campaign”. *BMJ*. <http://www.bmj.com/open-data>, (accessed 2018-10-31).

Thomson Reuter 社²⁹⁾は、データの引用索引である Data Citation Index (DCI) を公開し、Web of Knowledge³⁰⁾に収録している論文と研究データのリンクや引用データの提供を開始した。2014 年には、California Digital Library (CDL)、PLOS、DataONE によってデータのメトリクスを開発する Making Data Count プロジェクトが開始された。2017 年からは Make Data Count プロジェクト³¹⁾として Crossref³²⁾の引用データを用いたメトリクスの開発に着手している（池内, 2018c）。2018 年 9 月には Google Dataset Search のベータ版³³⁾が公開され、研究データとデータの引用情報の検索やライセンス情報の表示が可能となった。

データ公開に関する課題の解決を目指すコミュニティを 2 件取り上げる。1 つは国際組織の RDA、もう 1 つは日本の研究データ利活用協議会（Research Data Utilization Forum, RDUF）である。いずれも参加者が自主的に課題、たとえば“メタデータ標準のカatalog作成”や“データの法的相互運用性”などを提案し、解決に向けた方策を探る組織である。

RDA は、「障壁なき研究データ共有」をスローガンとして 2013 年 3 月に結成された。生命科学、地球科学、化学、情報学、人文学、社会科学など多分野の研究者をはじめ、IT 技術者、図書館員、出版社、助成機関、政府組織など多彩なステークホルダーが参加しており、2018 年 10 月時点で、137 か国から約 7,300 人がメンバー登録している³⁴⁾。その特徴は、“参加者からのボトムアップの提案にもとづき、コミュニティ・コンセンサスをベースにして研究データ共有のポリシー、方式・制度、技術、基盤システムなどについて検討、合意形成、実践していくことを目指している”（村山, 2016）点にある。参加者の関心や課題によって Working Group (WG) や Interest Group (IG) が立ち上げられ、メーリングリストやオンライン会議で意見交換を行い、年に 2 回の総会を経て、18 か月後には成果物として提言が報告される。提言は法的効力をもたないが、データ公開に関する国際的なデファクトスタンダードとなりうるとされている（国際的動向を踏まえたオープンサイエンス検討会, 2015）。G7 科学技術大臣会合も、RDA と連携しながらポリシーの共有やインセンティブの検討を行うとしている（G7 Science and Technology Ministers, 2016）。

RDUF は、RDA と同様にデータ公開の課題を共有し、解決に向けた取り組みを行う場として 2016 年に設立された。その前身は、ジャパンリンクセンター（Japan Link Center, JaLC）が 2014 年 10 月から 2015 年 10 月まで実施した「研究データへの DOI 登録実験プロジェクト」に参加した研究データ担当者のコミュニティであり、研究会や報告会、メーリングリストによる情報交換を行っている³⁵⁾。2018 年 10 月現在、機関メンバーとして科学技術振興機構 (JST),

²⁹⁾ 現在は Clarivate analytics 社。

³⁰⁾ 現在は Web of Science。

³¹⁾ Make Data Count (MDC)。 <https://makedatacount.org>, (accessed 2018-10-31).

³²⁾ Crossref. <https://www.crossref.org>, (accessed 2018-10-31).

³³⁾ Google Dataset Search. <https://toolbox.google.com/datasetsearch>, (accessed 2018-10-31).

³⁴⁾ “About RDA”. Research Data Alliance. <https://www.rd-alliance.org/about-rda>, (accessed 2018-11-30).

³⁵⁾ “RDUF について”. 研究データ利活用協議会. <http://japanlinkcenter.org/rduf/about/index.html>, (accessed 2018-10-31).

物質・材料研究機構（NIMS）、国立情報学研究所（NII）、国立国会図書館（NDL）、情報通信研究機構（NICT）、千葉大学附属図書館／アカデミック・リンク・センターが参加しているほか、個人の参加も受け付けている。また、RDA の WG や IG にあたる小委員会を募集しており、さまざまな課題に対応するための指針やガイドラインの作成を目指している。

2.3.2 学術情報政策の動向

データ公開やオープンサイエンスの推進に関する政策について、国際動向と日本の動向をそれぞれ述べる。なお、国際動向のうち G7（G8）や OECD の政策には日本も合意している。

[1] 国際動向

データ公開やオープンサイエンス政策の国際動向として、表 2-4 に示す OECD、G7（G8）、英国、米国、そして欧州連合（European Union, EU）の主要な政策について述べる。

表 2-4 データ公開・オープンサイエンス政策（国際）

OECD	2007 年	OECD principles and guidelines for access to research data from public funding
	2015 年	<i>Making Open Science a Reality</i>
G7（G8）	2013 年	G8 Science Ministers' Statement
	2016 年	G7 Tsukuba Communiqué（G7 Science and Technology Ministers）
	2017 年	G7 Science Ministers' Communiqué
英国	2007 年	BBSRC Data Sharing Policy（2010 年改定） Wellcome Trust Policy on Data Management and Sharing（2010 年改定）
	2011 年	RCUK Common Principles on Data Policy（2015 年改定）
	2016 年	Concordat on Open Research Data
米国	2003 年	NIH Data Sharing Policy で DMP の提出を義務化
	2011 年	NSF Grants Proposal Guide で DMP の提出を義務化
	2013 年	OSTP Memorandum for the heads of executive departments and agencies
	2016 年	IWGOS（Interagency Working Group on Open Science）設置
EU	2013 年	Guidelines on Data Management in Horizon 2020（都度改訂）
	2015 年	EOSC（European Open Science Cloud）計画発表
	2016 年	EOSPP（European Open Science Policy Platform）設置
	2017 年	EOSC Declaration

OECD 加盟国の科学技術大臣は、2004 年に研究データのアクセスに関する協議を行い

(Arzberger et al., 2004), 社会における重要性をふまえた宣言 (OECD, 2004) を採択した³⁶⁾。そして 2007 年には公的資金による研究データへのアクセスを拡大する「OECD 原則」が公開された (OECD, 2007)。2015 年には, *Making Open Science a Reality* と題する報告書を公開して, オープンサイエンスの状況と課題に対する 13 の政策メッセージを提示した (OECD, 2015)。

2013 年に G8 科学大臣会合が科学研究データ公開の基本原則を含む共同声明 (G8 Science Ministers, 2013) を公開した。これを契機として, 各国・地域の政府組織や助成機関によってデータ公開が推進され, 義務化の動きも広がった (SPARC Europe & Digital Curation Centre [DCC], 2017; SPARC Europe & DCC, 2018)。2016 年の G7 科学技術大臣会合では, オープンサイエンスを推進することが合意され, オープンサイエンス作業部会が設置された (G7 Science and Technology Ministers, 2016)。さらに 2017 年の G7 科学大臣会合では, 作業部会の報告を受けて, オープンな研究エコシステムのためのインセンティブと, 研究データを利用するためのインフラについて検討することが確認されるなど (G7 Science Ministers, 2017), オープンサイエンスへの取り組みは着々と進展している。

米国では, 2003 年に国立衛生研究所 (National Institutes of Health, NIH) がデータ共有ポリシーを公開し (NIH, 2003), データマネジメントプラン (Data Management Plan, DMP) の提出を義務化した。DMP とは, 研究のために作成, 収集するデータの情報や, 公開, 保存方法などについて記載した上で研究助成金を申請する際に助成機関などに提出する書類である (Jones, 2011)。データ公開に関する政策やオープンサイエンス政策の文脈では, DMP の提出を義務づけることをデータ公開の義務化とみなす場合が多い。2011 年には NSF が Grant Proposal Guide で DMP の提出を義務づけた (National Science Foundation [NSF], 2011)。そして 2013 年に, 米国科学技術政策局 (Office of Science and Technology Policy, OSTP) は年間予算 1 億ドル以上の政府機関に対して, 政府助成による研究成果 (論文およびデータ) のパブリックアクセスを促進させるための計画を策定するよう指示した (OSTP, 2013)。これを受けて, エネルギー省 (Department of Energy, DOE) や米国航空宇宙局 (National Aeronautics and Space Administration, NASA) などの省庁や機関がパブリックアクセス計画を公開し, その中で DMP の提出を義務づけている (Valen & Blanchat, 2015)。2016 年にはオープンサイエンスの一層の推進をはかるため, 米国科学技術会議 (National Science and Technology Council, NSTC) の科学委員会 (Committee on Science, CoS) が, オープンサイエンスに関する省庁間作業部会 (Interagency Working Group on Open Science, IWGOS) を設置した (Sheehan, 2017)。

英国では, 2007 年にバイオテクノロジー・生物科学研究会議 (Biotechnology and Biological Sciences Research Council, BBSRC) やウェルカムトラスト (Wellcome Trust) がデータ公開に関するポリシーを公開した。その後, 2011 年に英国研究会議 (Research Councils UK, RCUK) がデータ方針の共通原則を公開したことを受けて, 工学・物理科学研究会議 (Engineering and Physical Sciences Research Council, EPSRC) など各研究会議や助成機関はデータ方針を策定,

³⁶⁾ OECD 加盟国 30 か国のほか, 中国, イスラエル, ロシア, 南アフリカ政府も採択した。

あるいは改訂した (DCC, n.d.)。2016 年には RCUK, Wellcome Trust, イングランド高等教育助成会議 (Higher Education Funding Council for England, HEFCE), 英国大学協会 (Universities UK, UUK) の 4 機関による協定が締結された (Higher Education Funding Council for England [HEFCE], Research Councils UK [RCUK], Universities UK [UUK], & Wellcome Trust, 2016)。

欧州では、オープンサイエンスは EU によるデジタル単一市場 (Digital Single Market) 戦略や Horizon 2020 の一部として位置づけられ、欧州オープンサイエンスクラウド (European Open Science Cloud, EOSC)³⁷⁾、オープンサイエンスポリシープラットフォーム (European Open Science Policy Platform, EOSPP)、オープンサイエンスモニター (2.1.4) などのプロジェクトが進行している。時系列順にみると、2013 年に EC が Horizon 2020 におけるデータマネジメントのガイドライン (EC, 2013) を公開し、数度にわたる改定を行っている³⁸⁾。2015 年には EOSC 計画が発表され、2017 年に EOSC 宣言 (EC, 2017) を公開している。SPARC Europe と英国の DCC は欧州のデータ公開とオープンサイエンス政策に関する調査分析の結果を継続的に公開している (SPARC Europe & DCC, 2017)。2018 年 1 月版の報告書によれば、EU の加盟国 28 か国のうち 10 か国 (ベルギー, キプロス, デンマーク, フィンランド, フランス, ドイツ, リトアニア, オランダ, ポルトガル, 英国), およびアイスランド, ノルウェー, セルビア, スイスが研究データに関する国のポリシーを策定している³⁹⁾ (SPARC Europe & DCC, 2018)。

[2] 日本の動向

日本においても 2013 年の G8 科学大臣会合共同声明 (G8 Science Ministers, 2013) などを契機としてデータ公開やオープンサイエンスに関する議論が重ねられてきた。表 2-5 に主要な政策文書を時系列に沿って示す。オープンサイエンスの関連文献やプロジェクト, 新聞報道などで参照されることが多い文書はゴシック体で記した。

³⁷⁾ EOSC は、欧州の全分野の研究者 170 万人と科学技術専門家 7,000 万人を対象としたデータ共有基盤として構想されている。

³⁸⁾ 2018 年 9 月現在, 最新版のタイトルには, FAIR が追加されている。“H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020 (ver. 3, 2016)”, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, (accessed 2018-10-31).

³⁹⁾ ポリシーは策定していないが, 活動を行っている国は 12 か国であった (オーストリア, ブルガリア, クロアチア, チェコ共和国, エストニア, ハンガリー, アイルランド, イタリア, ポーランド, スロベニア, スペイン, スウェーデン)。

表 2-5 データ公開・オープンサイエンス政策に関する主要な文書（日本）

2014 年 9 月	「報告：オープンデータに関する権利と義務：本格的なデータジャーナルに向けて」（日本学術会議情報学委員会国際サイエンスデータ分科会）
2015 年 3 月	「回答：科学研究における健全性の向上について」 （日本学術会議）※データ等の資料は原則 10 年保存する
2015 年 3 月	「我が国におけるオープンサイエンス推進のあり方について：サイエンスの新たな飛躍の時代の幕開け」 （内閣府 国際的動向を踏まえたオープンサイエンスに関する検討会）
2015 年 4 月	「わが国におけるデータシェアリングのあり方に関する提言」 （科学技術振興機構科学技術情報委員会）
2016 年 1 月	「オープンイノベーションに資するオープンサイエンスのあり方に関する提言」（日本学術会議）
2016 年 1 月	「第 5 期科学技術基本計画（2016.4-2021.3）」 （内閣府）※第 4 章でオープンサイエンスの推進に言及
2016 年 2 月	「戦略的創造研究推進事業におけるデータマネジメント実施方針」 （科学技術振興機構戦略研究推進部・研究プロジェクト推進部）
2016 年 2 月	「学術情報のオープン化の推進について（審議まとめ）」 （文部科学省科学技術・学術審議会学術分科会学術情報委員会）
2016 年 7 月	「オープンイノベーションに資するオープンサイエンスのあり方に関する提言」（日本学術会議）
2017 年 1 月	「総合政策特別委員会における第 5 期科学技術基本計画の実施状況のフォローアップ等に関する審議のとりまとめ」（文部科学省科学技術・学術審議会総合政策特別委員会） ※(4)オープンサイエンスの推進に関する取組の在り方
2017 年 4 月	「オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針」（科学技術振興機構 [JST]）
2017 年 12 月	「委託研究開発におけるデータマネジメントに関する運用ガイドライン」 （経済産業省）
2018 年 3 月	「NEDO プロジェクトにおけるデータマネジメントに係る基本方針」 （新エネルギー・産業技術総合開発機構 [NEDO]）
2018 年 3 月	「データマネジメントプランの提出について」 （日本医療研究開発機構）
2018 年 6 月	「統合イノベーション戦略」（内閣府） ※2 章知の源泉 (2)オープンサイエンスのためのデータ基盤の整備

データ公開やオープンサイエンス全般に関するものとして、内閣府の国際的動向を踏まえたオープンサイエンスに関する検討会（2015）による報告書（以下、「内閣府報告書」）、科学技術振興機構（2015）による「わが国におけるデータシェアリングのあり方に関する提言」、文部科学省科学技術・学術審議会学術分科会学術情報委員会（2016）による「学術情報のオープン化の推進について（審議まとめ）」、日本学術会議（2016）による「オープンイノベーションに資するオープンサイエンスのあり方に関する提言」などがあり、国外の動向や実現に向けた課題が示されている。そして第 5 期科学技術基本計画（2016～2020 年度）では、“国は、資金配分機関、大学等の研究機関、研究者と連携し、オープンサイエンスの推進体制を構築する”（内閣府, 2016, p. 32）と述べられている。第 5 期科学技術基本計画に対して、文部科学省科学技術・学術審議会総合政策特別委員会（2017）はデータ共有・公開を推進するための DMP の導入、リポジトリなどの基盤の整備、評価など取り組みの方向性を示した。2018 年 6 月に閣議決定された「統合イノベーション戦略」では、「第 2 章知の源泉」に「(2)オープンサイエンスのためのデータ基盤の整備」が掲げられ、主要施策として“研究分野の特性等を踏まえたオープン・アンド・クローズ戦略を考慮したデータポリシーやデータマネジメントプランの策定を促進し、これらに基づく研究データの管理・公開等を促進するとともに、公的資金による研究成果としての研究データについては、データインフラを通して機械判読可能化を促進する”（内閣府, 2018, p. 16–19）とされている。

助成機関によるデータ公開の義務化、すなわち DMP の提出要求は、2016 年 2 月に科学技術振興機構戦略研究推進部・研究プロジェクト推進部（2016）が公表した「戦略的創造研究推進事業におけるデータマネジメント実施方針」と「運用ガイドライン」を嚆矢とする。米国、英国、EU では、一般に助成金の申請時に DMP の提出を義務付けているのに対して、この方針は一部の助成金に採択後の DMP 提出を求めるものであり、データは非公開とすることも可能であるという比較的緩やかな指示であった。この改訂版として 2017 年 4 月に「オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針」と「運用ガイドライン」が公開された（科学技術振興機構, 2017）。研究成果論文のエビデンスとなる研究データは公開を推奨され、研究代表者等は遅くとも研究開始までに DMP を提出するよう求めている。次いで 2017 年 12 月、経済産業省は「委託研究開発における知的財産マネジメントに関する運用ガイドライン」の別冊として、「委託研究開発におけるデータマネジメントに関する運用ガイドライン」（経済産業省, 2017b）を策定した。その内容は、2018 年 4 月以降に公募を開始する全てのプロジェクトを対象として、データ方針の作成や DMP の提出を求めるものである。これを受けて 2018 年 3 月、新エネルギー・産業技術総合開発機構（NEDO）（2018）は、「NEDO プロジェクトにおけるデータマネジメントに係る基本方針」を公開した。同じく 2018 年 3 月、日本医療研究開発機構（AMED）（2018）は「データマネジメントプランの提出について」を公開し、2018 年 5 月以降に新規公募する事業に対して DMP の提出を義務化すると発表した。

研究不正の側面からは、2014 年 8 月に文部科学省（2014）が公開した「研究活動における不正行為への対応等に関するガイドライン」に対して、2015 年 3 月に日本学術会議（2015）が「科学研究における健全性の向上について」を回答した。回答の中で、数値データを含む資料の“保存期間は、原則として、当該論文等の発表後 10 年間とする”こと、“電子化データについては、メタデータの整理・管理と適切なバックアップの作成により再利用可能な形で保存する”（日本学術会議, 2015, p. 8）こととしている。

これらの動きに先駆けて、2014 年 9 月、日本学術会議情報学委員会国際サイエンスデータ分科会（2014）は「オープンデータに関する権利と義務：本格的なデータジャーナルに向けて」と題する報告を行った。内容は、データ公開の目的やデータジャーナル、データによる分野連携などについて、CODATA 小委員会と WDS 小委員会の審議を取りまとめたものであった。

また、データ公開やオープンサイエンスに関わる動きとして、2018 年現在、著作権法（文部科学省, 2018）や不正競争防止法（経済産業省, 2017a）の改正案が議論されている。その目的の 1 つには、産学官を含む公開データの利用範囲を拡大することがあり、モノのインターネット（Internet of Things, IoT）や人工知能（Artificial Intelligence, AI）といった技術と組み合わせることによって、新たな知見や経済的な効果を生み出すことが期待されている。

2.3.3 図書館等による研究データ管理サービス

[1] 国際動向

2010 年ごろからデータ公開を支援するための研究データ管理（Research Data Management, RDM）サービスや研究データサービス（Research Data Services）を提供する大学・研究図書館が徐々に増加している（池内, 2014a; 池内, 2014b）。米国（Fearon, Gunia, Lake, Pralle, & Sallans, 2013; Tenopir, Sandusky, Allard, & Birch, 2014）や欧州（Cox & Pinfield, 2014; Tenopir, et al., 2017）の主要大学（Si, Xing, Zhuang, Hua, & Zhou, 2015）のみならず、小規模大学（Shorish, 2012; Clement, Blau, Abbaspour, & Gandour-Rood, 2017）やその他の地域（Primary Research Group, 2013; Kahn, Higgs, Davidson, & Jones, 2014; Renwick, Winter, & Gill, 2017）においても多数の事例が報告されている。

データの管理や公開を支援するデータライブラリアンについて、事例研究（Wang & Fong, 2015; Cox & Verbaan, 2016）や職能の検討（Davidson, 2014; Brown, 2015; Rice & Southall, 2016; Schmidt & Shearer, 2016b; Ravez, 2018）が行われてきた。研究データ管理に際しては、機関の研究支援部署や IT 部署、外部組織などと連携して支援体制を構築することが望ましいと指摘されている（Tenopir, et al., 2017）。Federer（2018）は、データライブラリアンを多様な分野を対象としたジェネラリストと分野に特化したスペシャリストに分類している。また、Lafferty-Hess, Rudder, Downey, Ivey, and Darragh（2018）はデータキュレーションに関するさまざまな

業務の難易度を、レベル 1（永続識別子の付与など）から、レベル 2（フォーマット変換など）、レベル 3（データクリーニングなど）まで 3 つのレベルに分類している。

[2] 日本の動向

日本では、RDM サービスの構築に向けた取り組みとして事例調査、RDM 教材の公開、実務のための検討や実践が進められている。事例調査の例として、デジタルリポジトリ連合（Digital Repository Federation, DRF）⁴⁰⁾による研究者へのインタビュー調査結果（デジタルリポジトリ連合, 2015; デジタルリポジトリ連合, 2016）、米国の訪問調査報告（安原, 小野, 2016）などが公開されている。

RDM 教材として、機関リポジトリ推進委員会（Institutional Repositories Promotion Committee, IRPC）⁴¹⁾による NISO の入門書『研究データ管理』の翻訳版（Strasser, 2016）が公開されている。2017 年 6 月に、オープンアクセスリポジトリ推進協会（Japan Confederation of Open Access Repositories, JPCOAR）は研究データタスクフォースを組織し、研究データ管理（Research Data Management, RDM）のスキルを習得するためのトレーニングツール⁴²⁾を公開した（常川ら, 2017）。そして同年 11 月、オープンサイエンス基盤研究センター（Research Center for Open Science and Data Platform, RCOS）⁴³⁾は、JMOOC/gacco のプラットフォームでオンライン講座「オープンサイエンス時代の研究データ管理」⁴⁴⁾を開講した。講座は、オープンサイエンスにおける研究データ管理に関する基礎的な知識の習得を目的として、大学・研究機関の図書館、IT 部門、研究支援部門等の支援人材育成を対象としたトレーニングツールとして開発された（古川, 2018）。

RDM の実践に向けた取り組みとして、2014 年月から 2015 年 9 月にかけて、JaLC による研究データへの DOI 登録実験プロジェクトが実施された（武田, 村山, 中島, 2015）。その成果として、「研究データへの DOI 登録ガイドライン」が公開された（ジャパンリンクセンター運営委員会, 2015）。また、機関リポジトリにデータを登録するためのメタデータ標準の検討も行われている（大園, 2017）。

⁴⁰⁾ 2016 年度末に解散した。

⁴¹⁾ 2017 年度よりオープンアクセス推進協会（JPCOAR）に移行した。

⁴²⁾ RDM トレーニングツール. <http://id.nii.ac.jp/1458/00000023/>, (accessed 2018-10-30).

⁴³⁾ 2017 年 4 月、オープンサイエンスのインフラとなる学術基盤を開発・運営することを目的として国立情報学研究所（NII）に設立された。 <https://rcos.nii.ac.jp>, (accessed 2018-10-30).

⁴⁴⁾ オープンサイエンス時代の研究データ管理. <https://www.nii.ac.jp/service/jmooc/rdm/>, (accessed 2018-10-30).

2.4 データ公開に関する実態調査

本節では、研究者によるデータ公開の実態や認識に関する先行研究の結果について述べる。なお、先行研究では「共有 (sharing)」という表現で本研究における「公開」や「提供」を示している場合があった。本節では、「公開」の意味を含む場合は原文通りに「共有」と記し、「提供」のみを扱う文献は取り上げないこととした。

2.4.1 データ公開に関する質問紙調査

研究者を対象としたデータ公開の実態に関する質問紙調査は、2000 年ごろから多数実施されてきた。しかし、2009～10 年と 2013～14 年の 2 度にわたって調査を実施した Tenopir et al. (2015) によれば、データ公開に影響を与える要因は経年的に変化している。また、データ公開に関する技術やネットワーク環境も経年的に変化している (Royal Society, 2012) ことから、本節では 2010 年以降に実施された、ある程度の規模をもつ調査を対象とする。まず、[1] 先行研究の概要を示した後、[2]データの公開率、[3]データの公開方法について述べる。

[1] 調査の概要

データ公開の実態や認識に関する質問紙調査を 12 件取り上げ、調査の概要として調査時期や対象について述べる。各調査が比較できるよう、表 2-6 に概要をまとめた。DataONE など「調査主体」の詳細は、調査内容とあわせて後述する。特定の国の調査の場合は「主な回答者」に国名を記載し、国際調査の場合は記載しなかった。日本の研究者の「回答数」や「公開率」⁴⁵⁾が明らかになっている場合は括弧内に人数を示した。

⁴⁵⁾ 公開率について、小数点以下まで示している調査はわずかであり、再計算もできなかったため、整数に統一した。

表 2-6 データ公開に関する質問紙調査の概要

参考文献／調査主体	調査年	主な回答者	回答数	公開率
Tenopir et al. (2011) /DataONE	2009 /2010	環境科学, 生態学等の研究者	1,329	36%
Huang et al. (2012)	2011	生物多様性分野の研究者	372	85%
Kim and Burns (2015)	2012 /2013	米国の生物学分野の研究者	608	
Kim and Stanton (2015)	2012 /2013	米国の STEM 分野の研究者	1,317	
Kim and Zhang (2015)	2012 /2013	米国の STEM 分野の研究者	1,298	
Damvad (2014) /RCN	2013	ノルウェーの社会科学, 健康科学, 自然科学等の研究者	1,474	28%
Tenopir et al. (2015) /DataONE	2013 /2014	環境科学, 生態学等の研究者	1,015	
Kratz and Strasser (2015a)	2014	生物学や考古学等の研究者	249	68%
Ferguson (2014) /Wiley 社	2014	多分野の研究者 (比率不明)	2,250+ (不明)	52% (44%)
Schmidt et al. (2016a) /Belmont Forum	2014	地球科学・環境学等の研究者, 技術員	1,253 (26)	
Fecher et al. (2015)	2014	独国の自然科学, 社会科学等の研究者	1,564	
小野ら (2016)	2015	日本の地球科学・環境学等の研究者	38 (38)	8% (8%)
Berghmans et al. (2017) /CWTS と Elsevier 社	2016	多分野の研究者	1,162	66%

※回答数や公開率に日本の研究者が含まれる場合は、括弧内に人数や比率 (%) を記した。

幅広い分野の研究者を対象とした先行研究として、まず、学術出版社である Wiley 社による調査を挙げる (Ferguson, 2014)。Wiley 社の調査は 2014 年に実施され、健康科学、生命科学、自然科学 (physical sciences)、人文・社会科学分野の 2,250 名を超える研究者から得た回答を分析した。国別の結果では、米国、英国、日本、中国、ブラジル、豪州、ドイツが比較されているが、回答者数の内訳は不明である。また、Damvad 社はノルウェー研究協議会

(Research Council of Norway, RCN) のため、2013 年に同国の研究者を対象とした調査を実施した。回答者は、社会科学、健康科学、自然科学など 1,474 名の研究者であった (Damvad, 2014)。ライデン大学の科学技術研究センター (Centre for Science and Technology Studies, CWTS)⁴⁶⁾と学術出版社の Elsevier 社は 2016 年 6 月から 7 月にかけて国際調査を実施し、1,162 名から回答を得た⁴⁷⁾ (Berghmans et al., 2017)。

分野は比較的限られているが、規模が大きい調査を 5 件取り上げる。Tenopir et al. (2011) は、Data Observation Network for Earth (DataONE)⁴⁸⁾を中心に、2009 年 10 月から 2010 年 7 月にかけてデータ公開の実践と認識に関する調査を実施した。回答者は環境科学や生態学の研究者が中心であるが、分野別の分析も行った。地域は北米が 73%、欧州が 15%、アジア・オセアニアが 7%であった。2013 年 10 月から 2014 年 3 月に後続調査を行い、1,015 名の回答を得た (Tenopir et al. 2015)。Schmidt et al. (2016a) は、ベルモントフォーラム (Belmont Forum)⁴⁹⁾が 2014 年に地球環境分野の研究者や技術者を対象として実施した国際調査の結果を報告した。回答者は、80 か国・地域の 1,253 名であった。また、Kim and Zhang (2015) は、2012 年 11 月から 2013 年 2 月にかけて、CoS の研究者データベースに登録された米国の STEM (科学・技術・工学・数学)⁵⁰⁾分野の研究者を無作為抽出して、データ共有への認識を調査した (1.1.3 に示した Kim and Stanton (2015) の調査と同じ)。Kim and Burns (2015) は、米国の生物科学者 608 名のデータ共有に関する回答も分析した。調査規模が大きく、分野によるデータ公開の動機や障壁が明らかにされているため、それぞれ取り上げた。

調査規模や分野の範囲は広くないものの、データの公開率などが明らかにされている研究として、Huang et al. (2012)、Kratz and Strasser (2015a)、小野、小池、柴崎 (2016) についても取り上げた。小野ら (2016) は、2015 年に日本の地球環境情報統融合プログラム (DIAS-P) とグリーン・ネットワーク・オブ・エクセレンス環境情報分野 (GRENE-ei) プロジェクトのコミュニティを対象とした意識調査を実施して、地球科学・環境学、農学、医学・健康などの研究者 38 名の回答を得た。調査の規模は大きくないが、日本の研究者を対象としているため参考とした。

⁴⁶⁾ ライデン大学に設置された学際研究機関。毎年、計量書誌学的手法を用いた CWTS Leiden Ranking (世界大学ランキング) を公表している。https://www.cwts.nl, (accessed 2018-10-31).

⁴⁷⁾ 本研究の終了後に公開されたため調査方法の参考にはできなかったが、結果の比較を行うためにここで取り上げる。

⁴⁸⁾ 米国にある地球科学分野の研究データプラットフォーム。2009 年に NSF などの資金提供を受けて設立された DataNet の国際プロジェクトの 1 つ。データリポジトリの Dryad や California Digital Library (CDL) など、多数の学術機関や団体が参加している。https://www.dataone.org, (accessed 2018-10-31).

⁴⁹⁾ 2009 年に設立された、地球の環境変動に関する研究への支援を行う世界各国の研究助成機関と国際的な科学組織のグループ。日本は文部科学省と科学技術振興機構 (JST) が参加している。
http://www.belmontforum.org, (accessed 2018-10-31).

⁵⁰⁾ Science, Technology, Engineering and Mathematics (STEM)

[2] データの公開率

データの公開率について、Tenopir et al. (2011) の調査で「他者が簡単に自分のデータにアクセスできる」という質問に「強く同意する」または「同意する」を選んだ回答者は 36%であり、ノルウェーの報告書で「全ての人」または「研究者」が「データを入手できる」とした回答者は 28%であった (Damvad, 2014)。Wiley 社の調査で「データを一般に公開している」回答者は 52%であり (Ferguson, 2014)、Berghmans et al. (2017) の報告では、何らかの方法でデータを公開している回答者は 66%であった⁵¹⁾。

データ公開が一般的な分野やデータ公開に関心のあるコミュニティを対象とした調査では、相対的に公開率が高かった。たとえば、生物多様性分野の主要な 3 誌⁵²⁾の著者を対象とした Huang et al. (2012) の調査では、論文に関連したデータを「常に」、「しばしば」、「時に」共有している回答者の合計は 85%であり、Kratz and Strasser (2015a) の調査では、生物学や考古学などの研究者 249 名のうち 68%がデータ共有経験を有していた。

日本の研究者について、Wiley 社の調査では「データを一般に公開している」回答者は 44%であった (Ferguson, 2014)。一方、小野ら (2016) の調査ではデータを「任意のユーザーに公開している」回答者は 8%であり差がみられた。

[3] データの公開方法

データの公開方法について、各調査で回答率が高かった順に 3 件ずつ挙げると、Huang et al. (2012) の調査では論文の補足資料を通じて (through files supplementary to articles) (52%)、パブリックデータベース (38%)、機関のサイト (25%) であり、Wiley 社の調査では雑誌の補足資料 (as supplementary material in a journal) (67%)、個人・機関・プロジェクトのウェブページ (37%)、機関リポジトリ (26%) であった (Ferguson, 2014)。Kratz and Strasser (2015a) の調査では直接コンタクト (87%)、リポジトリ (54%)、雑誌の補足資料 (journal website)⁵³⁾ (37%) であり、Berghmans et al. (2017) の報告では出版物の付録や補足資料 (appendix or supplement to publication) (33%)、データジャーナルでの出版 (28%)、データリポジトリ (13%) であった。以上の結果から、リポジトリやパブリックデータベースによるデータ公開は一定数行われているものの、まだ主流とはいえないことがわかる。

[2]に示したデータの公開経験や頻度を尋ねる質問方法は、回答者によって解釈が異なる可能性が排除できない。このように具体的な公開方法を選択する方法ならば、回答者による齟

⁵¹⁾ 出版物の付録や補足資料、データジャーナルでのデータ出版、データリポジトリ、その他、いずれの方法でもデータを公開していない回答者が 34%であったことから 66%とした。

⁵²⁾ *Biodiversity and Conservation*, *Diversity and Distributions*, *Journal of Biogeography* の 3 誌。

⁵³⁾ 同じ調査の結果を報告した *Scientific Data* の記事では、journal website (as supplemental material) と記載している (Kratz & Strasser, 2015b)。

齟齬をきたし難いと考えられる。

2.4.2 データ公開と関連のある経験

先行研究では、データ公開と関連のある経験について相反する知見が示されてきた。一方は、研究者のデータ共有やデータ利用経験がデータ公開経験と関連するという結果、もう一方は、研究者はデータを利用したり提供を求めたりするが、自身のデータは公開しないという結果である。

個人や研究グループ内におけるデータ共有（提供・被提供）経験（Damvad, 2014; Berghmans et al., 2017）や、公開データの利用経験がデータ公開経験に影響を与えているとする結果が示されていた。Piwowar（2011）は遺伝子発現マイクロアレイデータを作成した論文の著者を対象として調査を行い、過去にデータを再利用した経験がある場合にデータを公開する傾向があることを明らかにした。ノルウェーの報告書では、他の研究者のデータを利用したいという要望がデータ公開の増加に繋がることが示唆されていた（Damvad, 2014）。

一方、Tenopir et al.（2011）の調査では、「簡単にアクセスできるならば他の研究者のデータセットを使いたい」という設問に「強く同意する」または「同意する」を選んだ回答者は 84%であったのに対して、データを公開している研究者は 36%であった。*Science* 誌の査読者を対象とした調査においても、回答者の 76%は同僚にデータ提供を求めたことがあるにもかかわらず、80%は自身のデータを整備するための資金がないとしていた（Science, 2011）。つまり、データ利用を希望したり提供を受けたりするだけで、自身は公開をためらう研究者が存在する可能性が示唆されていた。

2.4.3 データ公開の動機と障壁

図 1-1（研究者からみたデータ公開の構造）に示したデータ公開の動機や障壁は、調査によって尋ね方や選択肢が異なる。たとえば Wiley 社の調査によれば、データ共有の動機は、研究コミュニティで標準的に実践されているから（57%）、研究のインパクトやビジビリティの向上（55%）、公益（50%）などであった（Ferguson, 2014）。小野ら（2016）の調査では、データ提供の動機は、科学の発展に対する貢献（86.8%）、分野間連携の促進のため（63.2%）、公的資金による研究成果のため（44.7%）などであった。CWTS/Elsevier 社の調査では、研究データ共有への態度として質問を行った結果、自分の分野では研究データ共有を重視している（69%）、出版者に提供したので論文とともに公開されると思う（52%）などの回答率が高かった（Berghmans et al., 2017）。こうしたデータ公開の動機や、次ページに示すさまざまな障壁から主要なものを選択するために、上位の項目を表 2-7 に示した。[] は調査項目（選択肢など）の形で明示されていなかったが、調査の結果をふまえてデータ公開に対する認識として考察されていた内容を表す。

表 2-7 質問紙調査によるデータ公開の主な動機と障壁

参考文献／調査主体	主な動機	主な障壁
Tenopir et al. (2011) /DataONE	[引用が重要]	時間, 資金, 権利, 公開する場所
Huang et al. (2012)	雑誌のデータアーカイブポリシー	出版前の公開, 公開の困難さ
Kim and Burns (2015)	助成機関と雑誌の圧力, データリポジトリとメタデータ標準の可用性	
Kim and Stanton (2015)	分野の規範, 科学的利他性, 雑誌の圧力, キャリアベネフィット	公開のための労力や時間についての認識
Kim and Zhang (2015)	キャリアベネフィット, [図書館による管理]	キャリアリスク, 公開のための労力
Damvad (2014) /RCN	[研究の進展, 研究の倫理的な義務]	時間, インフラ, インセンティブ, 将来の出版可能性
Tenopir et al. (2015) /DataONE	[引用が重要]	出版前の公開, 誤解や誤用
Ferguson (2014) /Wiley 社	研究コミュニティ, 研究のインパクト, 公益	知的財産権・機密, 所属機関・助成機関の要求がない
Schmidt et al. (2016a) /Belmont Forum	科学研究と成果実装の加速, 研究成果の認知向上	出版前の公開, 法的制約, 信用や正当な評価の喪失
Fecher et al. (2015)	評判経済 (評価やデータ引用)	他の研究者が先に出版すること
小野ら (2016)	科学の発展に貢献, 分野間連携の促進	時間, 提供作業, 自分が優先的に使う
Berghmans et al. (2017) /CWTS と Elsevier 社	[分野で重視, 論文とともに公開]	[訓練不足, 功績にならない]

[] は調査項目ではないが, データ公開に対する認識として考察されていた内容を示す。

データ公開の動機と障壁は, 表 2-7 に示した質問紙調査の結果以外にも, インタビュー調査などの方法によって明らかにされている。そこで, 研究者がデータを公開する動機と障壁を明らかにしている最近の調査, すなわち Wallis, Rolando, and Borgman (2013) や van den Eynden et al. (2014) の調査, Fecher et al. (2015) のシステマティックレビュー, Nature news チームのインタビュー調査 (van Noorden, 2014), Wicherts, Bakker, Molenaar, and Dylan (2011) による公開データの分析を参照して, 複数の先行研究で指摘されていた項目を表 2-8 に示す。

以上の先行研究による結果を総括して、表 2-8 に研究者の立場からみた主なデータ公開の動機と障壁を示す。動機は、学術雑誌や助成機関によるデータ公開ポリシーなどの制度（以下、「制度」）、および“業績になるから”、“科学の発展に貢献したいから”といった誘引（以下、「誘引」）にわけて整理した。障壁は、“公開するためのリポジトリ、資源、時間がない”といった資源の不足感（以下、「資源の不足感」）、および“先に論文を出版される可能性”、“誤解や誤用される可能性”といったデータ公開に対する懸念（以下、「懸念」）にわけて整理した。それぞれの出典は付表 1 に示す。

表 2-8 研究者からみたデータ公開の動機と障壁

動機	障壁
制度 <ul style="list-style-type: none"> ・ 学術雑誌のポリシー ・ 助成機関のポリシー ・ 所属機関のポリシー ・ 分野の規範 誘因 <ul style="list-style-type: none"> ・ 業績 ・ 科学の発展 ・ 他の研究者からの要求 ・ 研究成果の認知度向上 ・ オープンデータへの貢献 	資源の不足感 <ul style="list-style-type: none"> ・ リポジトリ ・ 資金 ・ 時間 ・ 人材 懸念 <ul style="list-style-type: none"> ・ 先に論文を出版される可能性 ・ 誤解や誤用の可能性 ・ 引用せずに利用される可能性 ・ 研究の誤りを発見される可能性 ・ 機密・プライバシー情報 ・ 商用利用される可能性 ・ 知的財産権

出典は付表 1 に示す。

各要因について、類似の内容は統合した。たとえば、Schmidt et al. (2016a) はデータ公開の障壁として、「誤解や誤用の可能性」と「データが独り歩きすることへの懸念」を挙げているが、後者はデータが文脈を無視して誤った解釈で利用されることへの懸念であり、広義には誤用に対する懸念であると考えて「誤解や誤用の可能性」とした。同様に、Tenopir et al. (2011) による「データがコントロールできなくなる恐れ」やノルウェーの報告書の「研究者がデータを理解できないことへの懸念」も「誤解や誤用の可能性」とした。また、Schmidt et al. (2016a) の「信用や正当な評価を失う恐れ」や Wicherts et al. (2011) の「再分析によってエラーや異なる結論が示される恐れ」は「研究の誤りを発見される可能性」とした。キャリアベネフィット、評価、引用などは「業績」、支援、労力などは「人材」とした。

「障壁」のうち「資源の不足感」は、研究者が不十分であると考えているという結果に基づく項目である。つまり、実際の資源が不足しているかどうかではなく、あくまでも回答者

の主観に基づく要因である。なお、資源が提供されるならば積極的なデータ公開の契機となるという結果や、データを公開している研究者は充足度が高いといった結果も示されていた⁵⁴⁾。また、研究者は知的財産権について懸念を抱いているという複数の調査結果がみられたが (Tenopir et al., 2011; Ferguson, 2014; Fecher et al., 2015; Schmidt et al., 2016a), 多くの国でデータやデータベースは知的財産権の対象として認められていない (OECD, 2015)。

表 2-7 と、表 2-7 を集約した表 2-8 から、データの公開の動機や障壁は多岐にわたること、多数の先行研究で選択されているような決定的な公開の動機や障壁はないことがわかる。また、表 2-8 に示した公開動機と障壁の程度は、年齢 (Tenopir et al., 2011; Tenopir et al., 2015; Schmidt, Gemeinholzer, & Birgit, 2016a) や分野 (Tenopir et al., 2011; Damvad, 2014; Ferguson, 2014; Tenopir et al., 2015; Schmidt et al., 2016a) などの属性によって差があることが明らかにされていた。たとえば年齢による違いについて、Schmidt et al. (2016a) の調査では、オープンデータを世に出す際の障壁として挙げられた 10 項目のうち、「データを公開する前に成果を出版したいという願望 (表 2-8 では「先取権の喪失」)」が最も重大な障壁であると認識されており、回答者の年齢が若いほどその傾向が強かった。Tenopir et al. (2015) の調査でも、11 項目の障壁のうち、「先に出版する必要がある」の選択率が最も高く (43.5%), かつ、最も若いグループの選択率が高かった。分野による違いについて、Wiley 社の調査では、健康科学の研究者のデータ公開率は 48% でありプライバシーや倫理的な懸念があるが、公益に資することが将来的なデータ共有の動機になること、その一方で人文・社会科学のデータ公開率は 36% であり、研究のインパクトやビジビリティを高める場合や助成機関の要求がデータ共有の動機になることなどを報告している (Ferguson, 2014)。

2.1.4 でも述べたように、G7 科学技術大臣会合は、“幅広い分野の公的資金による研究成果 (論文や関連するデータセット等) に学術関係者だけでなく、民間企業や一般市民が、広く利用・アクセスできるようにするものである” (G7 Science and Technology Ministers, 2016, p. 9) としている。総務省の統計によれば、調査時点における日本の研究者の約 6 割は企業の研究者であり⁵⁵⁾、大学と連携した産学共同研究も推進されている (総務省, 2016)。しかし、質問紙調査のうち、企業の研究者が明示的に含まれていたのは Tenopir et al. (2011) と Schmidt et al. (2016a) の 2 件のみであった。企業の回答者は、Tenopir et al. (2011) の調査では 34 名 (全体の 2.6%), Schmidt et al. (2016a) の調査では 32 名 (同 2.6%) であり、比率が少ないため、大学の研究者との比較は行われていなかった。

⁵⁴⁾ たとえば表 2-7 の Kim and Burns (2015) など。

⁵⁵⁾ 2016 年 3 月 31 日時点の日本の研究者数は 847,100 名、うち大学は 322,100 人 (38.0%), 企業は 486,200 人 (57.4%), 非営利団体・公的機関は 38,800 人 (4.6%) であった。

2.5 学術雑誌のデータ公開ポリシー

2.5.1 データ公開ポリシーの掲載率と強度の分類

[1] データ公開ポリシーの掲載率

学術雑誌によるデータ公開ポリシー（以下、「ポリシー」）に関する調査は、雑誌の投稿規定などを対象として行われてきた。ポリシーの調査内容を確認すると、データをリポジトリに登録する方法と補足資料に掲載する方法を区別する調査と、区別しない調査がみられた。また、ポリシーの有無のみを判断する調査と要求の強度を数段階に分類する調査がみられた。

表 2-9 に主な先行研究の概要として、著者名、調査年（不明の場合は論文の発表年）、調査対象、調査対象誌数、ポリシーの掲載誌数と掲載率、補足資料の調査の有無を示す。経年変化を確認するため、排列は調査年が古い順とした。なお、後述するように要求の強度の分類は調査によって異なるため、掲載率は強度にかかわらずポリシーの有無に基づく掲載率を示した。

表 2-9 先行研究による学術雑誌のデータ公開ポリシーの掲載率

著者	調査年	調査対象	対象誌数	掲載誌数	補足
McCain(1995)	1992*	自然科学, 医学, 工学	約 850	132 (15.5%)	○
Piwowar & Chapman (2008)	2007	遺伝子発現解析に関する 論文の掲載誌: 生化学・分 子生物学, 細胞生物学など	70	42 (60.0%)	
Alsheikh-Ali et al. (2011)	2009	IF 上位 50 誌: 生命科学, 複合領域など	50	44 (88.0%)	
Gherghina & Katsanidou (2013)	2011	政治学	120	19 (15.8%)	
Stodden et al. (2013)	2011	計算科学: 確率統計, 数理・	170	56 (32.9%)	○
	2012	計算生物学等	170	64 (37.6%)	
Vlaeminck (2013)	2013†	経済学 (ドイツ, スイス, オーストリアの雑誌)	141	29 (20.6%)	
Zenk-Möltgen & Lepthien (2014)	2014†	社会学	140	101(72.1%)	
Aleixandre-Benavent et al. (2016)	2014	情報学・図書館学	85	57 (67.1%)	○

*調査対象誌の出版年 †論文の発表年 「補足」=補足資料の調査の有無

McCain (1995) は、1992 年に出版された自然科学分野の約 850 誌を対象として、ポリシーを調査した。ポリシーの強度は、遵守しなかった場合のペナルティを「論文を出版しない (宣言または強く示唆)」、「論文を出版しない (控えめな表現)」、「明記なし」、「今後の出版物へのアクセスを拒否 (denied future publication access)」の 4 段階に分類している。調査の結果、132 誌 (15.5%) が研究関連情報 (Research-Related Information, RRI) ⁵⁶⁾ の公開や補足資料への掲載を求めるポリシーを有していた。

2007 年に Piwowar and Chapman (2008) はマイクロアレイによる遺伝子発現解析に関する論文の掲載誌 70 誌を対象として、ポリシーを調査した。ポリシーの強度は「強い」、「弱い」、「なし」に分類している。調査の結果、42 誌 (60.0%) が何らかのポリシーを掲載していた。

2009 年に Alsheikh-Ali, Qureshi, Al-Mallah, and Ioannidis (2011) は、Journal Citation Reports (JCR) を用いてインパクトファクター (Impact Factor, IF) 上位 50 誌を対象とした調査を行った。調査対象誌の分野は生命科学や複合領域が中心であった。調査の結果、44 誌 (88.0%) がマイクロアレイや核酸配列データのリポジトリ登録に関するポリシーを掲載していた。

2011 年に Gherghina and Katsanidou (2013) は、Citation Index 2010 にランク付けされた政治学分野の雑誌から国際誌かつ査読誌である 120 誌を選択して、ウェブサイトでポリシーの調査を行った。ウェブサイトにポリシーが掲載されていた 18 誌を除く 102 誌に電子メールで質問したところ、45 誌 (44.1%) から回答を得た。うち、1 誌はデータ公開ポリシーを有していたため、合計 19 誌 (15.8%) がポリシーを有していることがわかった。

2011 年と 2012 年に Stodden, Guo, and Ma (2013) は、確率統計、数理・計算生物学など計算科学分野 170 誌を対象として、データ、コード、補足資料の公開に関するポリシーを調査した。ポリシーの強度は「要求 (査読・出版条件として要求)」、「要求」、「推奨」、「言及のみ (implied)」、「記載なし」の 5 段階に分類している。調査の結果、データのポリシーを掲載する雑誌は一年間で 56 誌 (32.9%) から 64 誌 (37.6%) に増加していた。

2013 年に Vlaeminck (2013) が発表した論文では、ドイツ、オーストリア、スイスで出版された主要な経済学分野の雑誌 141 誌を対象として、データの可用性 (availability) ポリシーと再現ポリシーを調査した結果を報告している。調査の結果、29 誌 (20.6%) にデータの可用性ポリシーが示されていた。

2014 年に Zenk-Möltgen and Lepthien (2014) が発表した論文では、Social Sciences Citation Index (SSCI) に収録された社会学分野の雑誌 140 誌を対象としてポリシーの調査を行った結果を報告している。ポリシーの強度は、「あり」、「出版社や機関が推奨」、「なし」の 3 段階に分類している。調査の結果、101 誌 (72.1%) にポリシーが掲載されていた。

2014 年に Aleixandre-Benavent, Moreno-Solano, Sapena, and Sánchez Pérez (2016) は、Web of Science (WoS) の情報学・図書館学 (Information Science and Library Science, ISLIS) に収録さ

⁵⁶⁾ 配列データ、結晶データ、生物試料、大規模な表やテキスト、ソフトウェアなど。

れている 85 誌を対象として、補完的な資料（complementary material）⁵⁷⁾、再利用、リポジトリへの登録などのポリシーを調査した。調査の結果、57 誌（67.1%）が主題（thematic）リポジトリまたは機関リポジトリへの登録ポリシーを掲載していた。

以上の先行研究によって、自然科学および社会科学分野のポリシーの状況が明らかにされてきた。Stodden et al. (2013) の結果では、調査年が新しいほどポリシーの掲載率がやや高かったが、全体的には調査年が新しいほど掲載率が高いとまではいえず、分野による差がみられた。たとえば、Piwowar and Chapman (2008) や Alsheikh-Ali et al. (2011) が調査した生命科学の関連分野は掲載率が高く、計算科学や経済学は低かった。また、1.3.1 で述べたように補足資料はデータ公開の手段として重要であると考えられているが、調査を行っていない研究もみられる。従って分野間の比較を行うためには、先行研究の結果を用いるのではなく、同時期に分野横断でポリシーを調査する必要があると考えられる。

[2] データ公開ポリシーの強度

表 2-10 に、ポリシーの強度を分類した調査、およびポリシーの強度を分類した上で実際のデータ公開率と比較した Vines et al. (2013b) による調査（1.3.2）で用いられたポリシーの強度の分類を示す。いずれの研究も、ポリシーの表現を参考に強度の分類を行っていた。Stodden et al. (2013) は最も詳細な 5 段階に、McCain (1995) は 4 段階に、Piwowar and Chapman (2008)、Zenk-Möltgen and Lepthien (2014)、Vines et al. (2013b) は 3 段階に分類していた。

表 2-10 先行研究によるデータ公開ポリシーの強度の分類

著者	種数	表現
Stodden et al. (2013)	5	要求（査読・出版条件として要求）、要求、推奨、言及、記載なし
McCain (1995)	4	論文を出版しない（宣言または強く示唆）、論文を出版しない（控えめな表現）、明記なし、今後の出版物へのアクセスを拒否
Piwowar & Chapman (2008)	3	強い、弱い、なし
Zenk-Möltgen & Lepthien (2014)	3	あり、出版社や機関が推奨、なし
Vines et al. (2013b)	3	義務（mandatory）、推奨、なし

⁵⁷⁾ Aleixandre-Benavent et al. (2016) は、雑誌の「追加（additional）・補完（complementary）・補足（supplementary）資料」をまとめて「補完的な資料（complementary material）」と記している。その内容は、論文とあわせて雑誌のウェブサイトに掲載されるデータであり、本研究や他の先行研究の「補足資料」と同様であった。

2.5.2 データ公開ポリシーと関連のある要因

先行研究では、ポリシーの掲載率や強度と関連する雑誌の特徴として、OA ステータス、出版者⁵⁸⁾、IF などが調査されていた。また、Piwowar and Chapman (2008) や Stodden et al. (2013) は、関連のある特徴を用いて重回帰分析を行い、ポリシーの掲載確率や強度の予測モデルを作成していた。

OA ステータスについて、Piwowar and Chapman (2008) の調査では OA の 4 誌は全て何らかのデータ公開ポリシーを掲載していた。多変量解析の結果、OA は IF とともにマイクロアレイデータの共有ポリシーの強度に影響を与えていた。Stodden et al. (2013) の調査では、OA 誌の方が非 OA 誌よりもデータ公開ポリシーの掲載率が高いという結果であった。

出版者は、商業出版社よりも学協会の方がポリシーの掲載率が高いという結果が示されていた。Piwowar and Chapman (2008) の調査では、学協会は 82%、商業出版社は 50%がポリシーを掲載していた。Stodden et al. (2013) の予測モデルでも学協会の方がポリシーを掲載する確率が高かった。

IF は、値が高いほどポリシーの掲載率や強度が高いという複数の調査結果がみられた。Piwowar and Chapman (2008) は、ポリシーの強度が高いほど IF の中央値が高いことを示していた。Stodden et al. (2013) の予測モデルでは、IF が高い雑誌ほどポリシーを掲載する確率が高かった。Zenk-Möltgen and Lepthien (2014) の調査でも、IF が付与されている 135 誌については、5 年 IF (付与されていない場合は 2013 年の IF⁵⁹⁾) の値が高いほどポリシーの掲載率が有意に高かった。

Stodden et al. (2013) の調査では、2011 年よりも 2012 年の方がポリシーの掲載率が高かった。また、類似の分野を対象とした Piwowar and Chapman (2008) と Alsheikh-Ali et al. (2011) の結果を比較すると、2007 年は 60.0%、2009 年は 88.0%であり、調査年が新しいほど掲載率が高いと推測される。

以上の結果を表 2-11 に示す。「OA 誌」の列は、Piwowar and Chapman (2008) と Stodden et al. (2013) の調査では、OA 誌の方が非 OA 誌よりもポリシーの掲載率が高いといった結果を示していたためそれぞれ「○」を付した。「出版者」の列は、Piwowar and Chapman (2008) も Stodden et al. (2013) も学協会の方が商業出版社よりも掲載率が高いといった結果を示していたため、それぞれ「学協会」と記した。同様に「IF」の列は、IF が高いほど掲載率が高い可能性が示唆されていた Piwowar and Chapman (2008)、Stodden et al. (2013)、Zenk-Möltgen and Lepthien (2014) に「○」を付した。「調査年と掲載率」の列は、新しいほど掲載率が高い可能性が示唆されていた Stodden et al. (2013) に「○」を、Piwowar and Chapman (2008) と

⁵⁸⁾ 先行研究および本論文では、出版物を刊行する団体として「商業出版社」と「学協会」を取り上げるため、その総称として「出版者 (publisher)」という語を用いる。

⁵⁹⁾ 通常の (2 年分の被引用回数をカウントした) IF。社会学分野の論文は引用されるまでに比較的時間がかかるため、5 年 IF を採用したと考えられる。

Alsheikh-Ali et al. (2011) は掲載率を記載した。

表 2-11 ポリシーの掲載率や強度と関連がある雑誌の特徴

著者	調査年	OA 誌	出版者	IF	調査年と 掲載率
Piwowar & Chapman (2008)	2007	○	学協会	○	(60.0%)
Alsheikh-Ali et al. (2011)	2009				(88.0%)
Stodden et al. (2013)	2011 2012	○	学協会	○	○
Zenk-Möltgen & Lepthien (2014)	2014			○*	

*5 年 IF

2.6 まとめ

本章では、本研究の対象であるデータやデータ公開に関するさまざまな定義や議論を紹介した後、本研究の背景としてデータ公開やオープンサイエンスの進展状況を概観した。また、本研究の先行研究として、研究者を対象としたデータ公開の実態に関する質問紙調査と雑誌のポリシー調査について述べた。

研究者を対象としたデータ公開の実態調査を比較検討した結果、データ公開の動機や障壁は多岐にわたり、共通する要因を特定するのは難しいことがわかった。データ公開に関連する経験については、調査によって相反する結果が示されていた。また、日本の研究者や、企業の研究者を対象とした調査と分析は充分に行われていなかった。

雑誌のポリシー調査の結果から、分野によってデータ公開要求の程度に差があることが示唆された。しかし未調査の分野が多く調査方法も先行研究によって異なることから、分野間比較を行おうとする場合には、新たに調査を実施する必要があると考えられる。

以上の点をふまえて、3 章では日本の研究者を対象とした質問紙調査を、4 章では多分野を対象とした雑誌のポリシー調査を行う。

3 章 日本の研究者によるデータ公開状況と内的要因

[RQ1]日本の研究者はどの程度データを公開しているのか、および[RQ2]日本の研究者によるデータ公開の内的要因は何かを明らかにするために、質問紙調査を実施した。具体的には2016年11月から12月にかけて、文部科学省科学技術・学術政策研究所（NISTEP）科学技術動向研究センターが運営する科学技術専門家ネットワーク（1,983名）を対象として、ウェブアンケートシステムを用いた調査を行った¹⁾。1,398名（有効回答率70.5%）の回答について、図1-2（比較分析の枠組み）に基づき分野別、所属機関別、年齢層別の分析を行い、論文のオープンアクセス（Open Access, OA）や先行研究の結果と比較した。本章は、3.1で調査・分析方法、3.2で結果について述べ、3.3で考察を行い、3.4にまとめを示す。

3.1 方法

まず予備調査を行い、質問紙の内容を決定した。本節では予備調査の概要、先行研究と予備調査をふまえた質問の作成方針、最終的な質問項目について述べる。続いて調査対象、調査方法と調査期間の詳細、集計・分析方法、回答率と回答者の属性について述べる。

3.1.1 予備調査

本調査に先立って、予備調査を3回実施した。予備調査の目的は、質問の順番やワーディングを検討することである。対象は本調査と同様に、大学や企業の研究者とした。第1回（2016年8月26日～30日）は紙で実施し、第2回（11月4日～7日）と第3回（11月21日～24日）は本調査と同じウェブアンケートシステムを用いて実施した。第1回は15名、第2回は8名、第3回は12名の協力を得た。

予備調査による主な改善点は次の2点である。(1)研究によって扱うデータの種類や量が異なるという指摘から、“論文などの成果物を出版済みの最新の主な研究1件のためのデータ”を「カレントデータ」と定義した上で、質問を行うこととした。(2)データ公開には馴染みの

¹⁾ 著者が客員研究官として所属するNISTEP科学技術予測センターに質問紙調査の企画を提案し、了承を得て実施した。本論文の核となる論文、すなわち本章に示す分析や考察の一部を述べた論文（池内, 2018a）の受理後に、集計結果や自由回答を報告書として公開した（池内, 林, 赤池, 2017）。報告書では、単純集計やクロス集計の結果のみを示している。報告書の分野別の集計には科学工学指標（Science and Engineering Indicators, S&EI）による13分類に「人文学」を追加した14分類を用いた。一方、本論文および核となる論文では、Essential Science Indicators（ESI）の22分類を用いた。

ない回答者が多いという指摘や、“論文もデータも公開していないと答え難い”という意見から、質問の順番を入れ替えるとともに、概要説明に“論文やデータの公開経験がない、というご回答も参考になりますので、ぜひ、率直なご意見をお聞かせ下さい。”という一文を追加した。このほか、“質問の意味が取りにくい”、“選択肢の差がわかりにくい”、といった指摘を受けた箇所についても適宜修正を行った。詳細は、それぞれの箇所で述べる。

3.1.2 質問の作成方針と項目

ここでは質問紙調査全体にかかわる方針について説明した上で、具体的な質問項目を示す。

まず、データ公開の経験がない回答者やなじみが薄い回答者が存在すると予想されたこと、データ公開の経験があっても回答者によって想定する内容が異なると予想されたことから、各セクションのはじめに 1.5.1 の定義と具体例を示した。また、こうした説明部分は読まれな可能性のあることから、経験の有無と方法を同一の質問で尋ねることによって、回答者ごとの認識の違いによる回答のずれを防ぐよう努めた。たとえばデータ公開の経験を尋ねる質問では、「学術機関のリポジトリ・データアーカイブ（大学や NASA のリポジトリなど）」というように具体的なデータ公開方法を列挙して、1 つ以上を選択した回答者はデータ公開経験が「ある」とみなし、あわせて排他的な選択肢として「いいえ」、「わからない」を提示してデータ公開経験の有無を判断することとした。質問の順番は、回答が比較的容易であると考えられる、論文の OA に関する質問を最初に配置した。

先行研究の結果と比較を行うためには同じ質問を行うことが望ましいが、全体の質問数と回答者の疲労効果（鈴木, 2016, p. 196）²⁾を抑制するために次の方針で作成した。(1)先行研究に似たような選択肢がある場合は、表 2-8 と同様に統合する³⁾、(2)尺度項目の一部は複数選択方式に変更して、尺度を選択する質問を合計 30 項目以下とする（Dörnyei, 2006）⁴⁾。また、無回答や適当な回答を避けるために、「わからない」という選択肢を加えた。

予備調査において、研究によって扱うデータの種類や量が異なるという指摘があった。そこでデータ公開の障壁に関する質問などについては「論文などの成果物を出版済みの最新の主な研究 1 件のためのデータ」を「カレントデータ」と定義した上で回答を求めた。また、データ公開経験が「ない」または「わからない」を選択した回答者にも公開することを想定してもらうため、例として「研究 1 件のために質問紙調査とインタビュー調査を実施した場合」を示した。この場合、対象となるカレントデータは、「質問紙の回答を入力したスプレッドシート、回答を分析するための R のコード、インタビューを録音した音声データ、イン

²⁾ “質問の後半にみられるいいかげんな回答や間違い”を指す。

³⁾ たとえば、「データが独り歩きすることへの懸念」（Schmidt et al., 2016a）、「データがコントロールできなくなる恐れ」（Tenopir et al., 2011）、「研究者がデータを理解できないことへの懸念」（Damvad, 2014）を「誤解や誤用の可能性」（Schmidt et al., 2016a）に統合した（2.4.3）。

⁴⁾ 最終的な尺度項目は、合計 19（7 問）である。

タビューを書き起こしたテキストデータなど」であり、データの公開とは、「質問紙調査の回答データを第三者が再利用できるように、調査概要の説明を作成、回答を入力したスプレッドシートから個人情報を削除、項目に見出しをつけるといった処理を行った上で、質問紙や分析のためのコードとともにリポジトリに登録する」とした。分野別の分析には、調査対象者が自己申告済みの Essential Science Indicators (ESI) の 22 分野を用いた。質問紙では、回答の負担を軽減するため科学工学指標 (Science and Engineering Indicators, S&EI) に「人文学」を追加した 14 分類 (1.5.2 参照) を提示した。

以上の方針に基づき作成した本調査の質問数は合計 35 問であり、回答を容易にするために類似の内容をまとめて 7 つのセクションに分割した。次ページの表 3-1 に質問項目一覧を示す。質問紙本文は [付録 2] に掲載する。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-1 質問項目一覧

質問	必須	方式
1. 研究分野について		
Q1. ご自身の研究分野に最も近いものをお選び下さい。	*	
2. 学術論文について		
Q2. ご自身の論文について、あてはまるものをお選び下さい。	*	
Q3. 論文をオープンアクセスにした理由として、あてはまるものをお選び下さい。	*	m
Q4. 論文がオープンアクセスではない理由として、あてはまるものをお選び下さい。	*	m
Q5. Q4の理由が解決された場合、論文をオープンアクセスにしたいと思われませんか？	*	
Q6. 研究に利用したことがある（アイデアの参考にしたり引用した経験がある）論文の分野をお選び下さい。	*	m
Q7. 論文を探す際に、よく利用する検索ツールや情報源をお選び下さい。		m
Q8. 論文を利用する際に、その信頼性の判断基準としている項目をお選び下さい。		m
3. 研究データの提供について		
Q9. 共同研究者を除く他の研究者にデータを提供したご経験はありますか？		4
Q10. 共同研究者を除く他の研究者からデータの提供を受けたご経験はありますか？		4
4. 公開データの利用について		
Q11. これまでに、公開データを以下の公開先から入手したご経験はありますか？	*	m
Q12. 公開データの入手にあたって、問題だと感じたことがある項目をお選び下さい。	*	m
Q13. 入手した公開データの利用目的について、あてはまるものをそれぞれお選び下さい。	*	3
Q14. 研究に利用したことがある（参考にしたり再分析した経験がある）公開データの分野をお選び下さい。	*	m
Q15. 公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい。		m
Q16. 公開データを利用する際に、信頼性の判断基準としている項目をお選び下さい。		m
Q17. 今後、利用してみたいと思う公開データの分野をお選び下さい。	*	m
5. 研究データの公開について		
Q18. これまでに、研究データを以下の方法で公開したご経験はありますか？	*	m
Q19. 研究データを公開した理由として、あてはまるものをお選び下さい。	*	m
Q20. 研究データを公開していない理由として、あてはまるものをお選び下さい。	*	m
Q21. Q20の理由が解決された場合、研究データを公開したいと思われませんか？	*	
Q22. 所属機関では、研究データの保存期間は定められていますか？		
6. 最近の研究のためのデータについて		
Q23. カレントデータの総量は、およそどれくらいでしたか？ 論文などには使用しなかったデータも含めてあてはまる単位をお選びください。		s
Q24. ご自身以外で、カレントデータの所有権をもつ人・組織をお選び下さい。		m
Q25. カレントデータには、以下の機密情報が含まれていますか？		m
Q26. カレントデータは何年くらい保存する必要があると思われるですか？ 年数を数字で入力して下さい。（保存の必要はないと考える場合は「0」、永久保存の場合は「999」と入力して下さい）		
Q27. 現在までに、カレントデータ（データの一部）を公開しましたか？	*	
Q28. カレントデータを管理・公開しようとする場合、次の資源は十分に整っていますか（いましたか）？	*	4
Q29. カレントデータを論文の発表前に公開しようとする場合、次の点は問題（懸案）となりますか（なりましたか）？	*	4
Q30. カレントデータを公開した場合（公開したカレントデータは）、ご自身と同じ分野、すなわち Q1 で選択した分野の研究者の多くが理解できると思われますか？	*	3
Q31. カレントデータを公開した場合（公開したカレントデータは）、異分野の研究者の多くが理解できると思われますか？	*	3
Q32. カレントデータを公開した場合（公開したカレントデータについて）、ご関心がある項目をお選び下さい。		m
Q33. カレントデータを整備・公開する上で、より詳しく知りたいと思われる項目をお選び下さい。		m
Q34. カレントデータの整備や公開を、ご自身や共同研究者にかかわって図書館員やデータキュレーターなどの第三者が行う場合、分野の知識や専門性が必要であると考えられる項目をお選び下さい。カレントデータが多様な場合は、もっとも難しいと考えられるデータについてお答え下さい。		m
7. 自由回答		
Q35. 論文やデータの公開、オープンサイエンス、および調査に関するご意見や感想がありましたら、ご自由にお書き下さい。		

複数選択法=m, スケール (3 件法=3, 4 件法=4, データサイズ=s)

3.1.3 研究課題と質問の設定

表 3-1 に示した質問内容を，分析を行うグループごとにまとめて表 3-2 に示す。

表 3-2 質問の構成

No	データ	質問内容	参考文献*	OA
[1]	Q18	データの公開経験	F, H, K1	Q2
[2]	Q19	データの公開理由	表 2-8	Q3
[3]	Q20	データの非公開理由	表 2-8, T	Q4
	Q21	データの公開意思		Q5
[4]	Q28	カレントデータの公開に必要な資源の状況	表 2-8, S, T	
	Q29	カレントデータ公開の懸念	表 2-8, S, T	
[5]	Q9	データの提供頻度		
	Q10	データの被提供頻度		
[6]	Q11	公開データの入手経験	F, H, K1	
	Q12	公開データの入手における問題	S	
	Q13	公開データの利用目的	K2, P	
	Q14	利用したことがある公開データの分野	T	Q6
	Q15	公開データの検索ツールと情報源	K2, S	Q7
	Q16	公開データの信頼性の判断基準	K2	Q8
	Q17	今後利用してみたい公開データの分野	T	
[7]	Q23	カレントデータの量		
	Q24	カレントデータの所有権	表 2-8, O	
	Q25	カレントデータの機密情報	O	
	Q30	カレントデータの理解（自分野）		
	Q31	カレントデータの理解（異分野）		
	Q22	所属機関のデータ保存期間規定の有無		
	Q26	カレントデータの望ましい保存期間		
	Q27	カレントデータ公開の有無	F, H, K1	
	Q32	メトリクスへの関心	K2	
	Q33	カレントデータの公開に関する知識・関心	J	
	Q34	カレントデータの管理・公開の専門性	J	
	Q35	自由回答		

*F=Ferguson(2014), H=Huang et al.(2012), J=Jones et al.(2013), K1=Kratz and Strasser(2015a), K2=Kratz and Strasser(2015b), O=OECD(2015), P=Piwowar(2011), S=Schmidt et al.(2016a), T=Tenopir et al.(2011), Tenopir et al.(2015)

前ページの表 3-2 について、「No」は分析を行うグループであり、次節では[1]から[7]に沿って結果を示す。「データ」は質問番号、「OA」はデータと同様に論文の公開経験や理由について尋ねた際の質問番号を表す。「参考文献」には、参考とした先行研究の著者の頭文字、および表 2-8 (研究者からみたデータ公開の動機と障壁) を参考とした場合は表番号を記した。表 3-2 のグレーの網掛け部分はデータ公開環境の整備や対策を検討するための質問であるため、本論文では扱わない。以下では、各質問項目の目的と設定方法について述べる。

[1] データの公開経験

日本の研究者によるデータ公開の状況を明らかにするとともに、OA の状況と比較するために、データ公開および OA 経験の有無を尋ねた。また、データを公開している回答者と公開していない回答者を判別して、以降の調査や分析に用いた。

Ferguson (2014), Huang et al. (2012), Kratz and Strasser (2015a) の調査を参考に、「所属機関のリポジトリ・アーカイブ」、「論文の補足資料 (supplementary materials)」など 7 項目をデータ公開の方法を選択肢として列挙した (Q18)。データ提供に関する選択肢 (著者に直接コンタクトなど) は削除した。選択肢の表現はできるだけ先行研究に即したが、予備調査で“わかりにくい”と指摘があった項目は、例を追記するなど表現を修正した。

[2] データの公開理由

データ公開と OA の主な理由を明らかにするために、それぞれの経験がある回答者を対象として公開理由を尋ねた。公開理由の選択肢は、表 2-8 に示した先行研究の「動機」(「誘因」と「制度」) とした (Q19)。ただし、動機のうち「業績になるから」について、論文の場合は OA かどうかにかかわらず、一般的に出版すること自体が業績になる (Fecher, Friesike, & Hebing, 2015) ため、OA の選択肢には含めなかった。

[3] データの非公開理由と公開意思

データ公開と OA の経験がない回答者を対象として、非公開の理由を尋ねた。非公開理由の選択肢は、表 2-8 に示した先行研究の「動機」がないこと (たとえば、「雑誌のポリシーがないこと」)、および「障壁」のうち「資源の不足感」があること (リポジトリ、資金、時間が不足していると認識していること) とした (Q20)。動機のうち「他の研究者からのリクエストに応じて」は、Tenopir et al. (2011) の選択肢を参考に「ニーズがないと思うから」とした。

また、動機が生じたり資源の不足感が満たされれば (たとえば、雑誌のポリシーが策定されたり、資源が整備・提供されれば) データ公開につながるのかどうかを確認するために、

非公開理由が解決した場合、データを公開する意思があるかどうかを尋ねた（Q21）。

[4] 障壁の程度

データ公開の障壁は、データを公開していない研究者のみならず、公開経験がある研究者にとっても問題であることが明らかにされていた。そこでデータ公開経験の有無にかかわらず、全ての回答者に「データを公開する場合」を想定して、先行研究で明らかにされたデータ公開の障壁がどの程度問題であるのかを尋ねた。具体的には、データ公開のための資源の充足度（Q28）と懸念（Q29）について、それぞれの重要度を尋ねた。予備調査の結果から、同じ研究者であっても研究によってデータの状況が異なると考えられたことから、回答者がどのデータについて回答すればよいのか特定できるよう、カレントデータについて回答を求めた。また、知的財産権についてはデータやデータベースが対象となっていないことについて。

障壁について、Tenopir et al. (2011) や Tenopir et al. (2015) は複数選択方式で尋ねていた。一方、Schmidt et al. (2016a) は「重大な障壁である」、「障壁である」、「軽微な障壁である」の3件法で尋ねていた。その結果、「重大な障壁である」の選択率が21%から54%まで差がみられたことから、尺度項目とすることが適していると判断した。また、先行研究では「知的財産権を失う可能性」が障壁として挙げられていたが、調査時点において日本ではデータベースに著作権が認められた例が数件あるのみであったため（末吉, 2012）、選択肢から除外した。

先行研究では論文の出版前にデータを公開することが想定されていた。しかし予備調査では、論文出版前のデータ公開は理解し難いという趣旨のコメントが複数みられたため、質問文に「カレントデータを論文の発表前に公開しようとする場合、次の点は問題（懸案）となりますか（なりましたか）？」と記載した（Q29）。

[5] データの提供と被提供状況

先行研究では、データの提供経験がデータ公開に繋がるという結果と（Damvad, 2014; Berghmans et al., 2017）、データの提供を受けるだけでデータを公開しない研究者がいるという結果が示されていた（Tenopir et al., 2011; Huang et al., 2012）。そこで、全ての回答者にデータの提供と被提供経験の両方を尋ねて、データ公開経験との関連を確認することとした（Q9, Q10）。

[6] 公開データの入手経験

[4]と同様に、先行研究では公開データの利用経験をデータ公開の動機とする結果と、データを公開していないが、公開データを利用したい研究者がいるという結果がみられた（Tenopir

3 章 日本の研究者によるデータ公開状況と内的要因

et al., 2011; Science, 2011)。そこで、全ての回答者に公開データの入手経験や研究への利用状況を尋ねて、データ公開経験との関連を確認することとした。公開データの入手方法は、[1] データ公開経験と同じ選択肢で尋ねた (Q11)。

データを入手しても研究には利用しない回答者が存在する可能性があるため、Kratz and Strasser (2015b) の調査を参考に、入手した公開データの利用目的を「自身の研究のアイデアや仮説の参考にする」、「再分析・再利用して自身の研究を行う」⁵⁾、「(他者の) 研究を再現・追試する」に分けて、それぞれの頻度を尋ねた (Q13)。また、入手したデータの信頼性はどのように判断しているのかを尋ねた (Q16)。

公開データ入手における問題は、Schmidt et al. (2016a) の「データにアクセスする際の負担感」(3 件法) を参考として、複数選択方式で尋ねた。選択肢には、「データの収集基準が異なる」と「データごとに品質が異なる」があったが、データの収集基準が結果としてデータの品質に繋がると考えられること、調査結果を確認すると両者の負担の度合いがほぼ同じであったことから前者のみを採用した (Q12)。公開データの利用意思は、漠然とした回答とならないように、既に研究に利用したことがある論文 (Q6) やデータ (Q14) の分野を尋ねた上で、今後データを利用してみたい分野について、複数選択方式で尋ねた (Q17)。

[7] データのプロファイル

「内閣府報告書」では、研究分野の特性に対する配慮として、“個々のデータセットの取り扱いに対する技術的容易さ、困難さなど、様々な固有の事情が各分野にありうることを認識しておく必要がある” (内閣府, 2015, p. 20) と指摘されている。そこで研究テーマや研究者によって扱うデータの種類が異なり、データ公開に影響しているという仮説を立てて調査した。具体的には、(1) データ量が多い (Q23)、(2) データの所有権が複数の人や組織にある (Q24)、(3) データに機密情報が含まれる (Q25)、(4) データの理解が難しいと考えている (Q30, Q31) 場合は、データを公開しない傾向にあるという仮説を立てた。カレントデータの機密情報の選択肢は OECD (2015) を参考に「個人情報」、「企業・商業上の機密情報」、「健康情報 (遺伝情報, 医療情報)」、「その他、守秘義務がある情報」とした。

3.1.4 調査対象

調査対象は、NISTEP 科学技術予測センターが運営している「科学技術専門家ネットワーク」⁶⁾とした。科学技術専門家ネットワークとは、科学技術に関する動向や見解などを収集す

⁵⁾ Kratz and Strasser (2015b) は、“論文の主な結論を支持する”と“論文の主な結論に到達する”という質問を行っていた。しかし、予備調査で「両者の違いが分かりにくい」との指摘が複数あり、詳しい説明を行うと煩雑になるため統合した。

⁶⁾ “科学技術専門家ネットワーク”, 文部科学省科学技術・学術政策研究所 (NISTEP)。

るための、産学官の研究者、技術者、マネージャーなど約 2,000 名の専門家集団である。メンバーは紹介によって選出される。調査時点での所属の構成比は大学 60.4%、企業 22.7%、公的機関・団体 16.0%、その他 0.9%であり、多分野かつ幅広い年齢層の研究者が含まれること、企業の研究者が含まれること、これまでの調査から 3〜4 割の回答率が見込めることから調査対象とした。

調査時点における日本の研究者 847,100 名の所属は大学 38.0%、企業 57.4%、非営利団体・公的機関 4.6%であり（総務省, 2016）、科学技術専門家ネットワークの構成比は実態とやや異なっている。また、科学技術専門家ネットワーク分野の構成比も工学分野の研究者の比率が高く（34.8%）、実態と差があると考えられる⁷⁾。従って、日本の全研究者から対象者を無作為抽出して調査を実施する方が、より精度の高い結果が得られるかもしれない。しかし、本調査はデータ公開や OA に馴染みがない研究者にとっては回答の負担が大きく、十分な回答が得られるかどうかかわからない。また、データ公開や OA の経験や関心がない研究者からは、あまり回答が得られない可能性がある。それよりも、科学技術に関する調査に慣れている科学技術専門家ネットワークの研究者に尋ねる方が、ネガティブな意見も含めて率直な回答が得られるのではないかと考えた。なお、本調査は 1.5.1 で定義した通り、「研究の成果である論文やスライドの根拠となる」データを対象としているため、研究分野を尋ねる Q1 の選択肢に「研究を行ったことはない（口頭発表や論文出版の経験はない）」を挙げて、当該回答者には自由回答のみ記入していただいた。

3.1.5 調査方法と期間

調査には、NISTEP が契約している Questant 社のウェブアンケートシステムを用いた。調査依頼文書や調査票を作成した上で、ウェブアンケートシステムの設定、調査依頼メールの送信は、NISTEP の担当者に依頼した。ウェブアンケートシステムは、多重回答を防ぐため回答者ごとに個別 URL が作成される。また、回答完了後には再度回答が行えないよう設定されている。アンケートのトップページには、回答者が事前に全ての質問を確認できるように PDF 版の調査票を掲載した。原則としてウェブアンケートシステムでの回答を求めたが、問題が生じた場合は Word ファイル等による提出も受け付けることとした。

調査期間は 2016 年 11 月 30 日から 12 月 9 日とした。11 月 30 日に、科学技術専門家ネットワークの 1,983 名を対象として、アンケートへの協力を依頼する電子メールを送信した。電子メールは NISTEP のシステムによって個人宛に送信され、本文に個別 URL が示されるため、他者による回答などの可能性は極めて低い。また、同一 URL から 2 回以上回答できない

<http://www.nistep.go.jp/activities/st-experts-network>, (accessed 2018-10-31).

⁷⁾ 総務省(2016)による、大学の「学問分野別研究本務者」は 11 分野に分類されており、うち工学分野の比率は 18.2%である (p. 40)。

ように設定している。リマインダは、未回答者を対象として12月2日と7日に送信した。その際には、回答率が低い分野を示して協力を仰いだ。12月10日以降も回答入力があったため、最終的に12月14日の回答までを結果に含めた。なお、回答のうち1件はWordファイルによる提出であったため、ウェブアンケートシステムに代行入力した上で⁸⁾、他の回答とあわせて出力した。

3.1.6 集計・分析方法

結果の集計に先立って、回答データのクリーニングを行った。選択肢「その他」に入力された記述のうち、適切な選択肢があると判断できる場合は、当該選択肢を選んだものとして修正した。また、「その他」や「自由回答」の記述で誤字・脱字と判断できる箇所は適宜修正した。尺度項目は、評定数値が大きいほどポジティブな回答に、小さいほどネガティブな回答に統一した。自由記述はTiny TextMiner (TTM) for Mac OSXを用いて頻出語を抽出した後、類義語の整理を行い、文脈から内容を分類してまとめた。自由回答は、報告書とあわせてcsvファイルとして公開している(池内, 林, 赤池, 2017)。

クリーニング後の回答は、ExcelとSPSS Version 24を用いて集計や分析を行った。分野や属性(年齢, 所属), およびデータを公開している研究者の特徴, データ公開に消極的な研究者の特徴を明らかにするために、各質問項目についてクロス集計表を作成した。年齢は, Schmidt et al. (2016a)の結果と比較するために、5年ごとに集計した。結果のうち、比率は小数第1位まで、相関係数などは小数第3位まで示した。先行研究もこれに準じたが、文献に示されておらず再計算もできない場合は原文のままとした。

データの提供経験と被提供経験の頻度など、変数同士に関連があるかどうかを比較する際、数値はPearsonの、順序尺度はSpearmanの相関係数を用いて関連を調べた。有意水準 α は5%として、有意である場合は相関係数 r や p 値を示した。分野別に順序尺度同士、または順序尺度と数値を比較する際には、順序尺度の選択率を適宜集計して数値化した上でPearsonの相関係数を用いた。たとえば、選択項目が「不十分」、「やや不十分」、「やや充分」、「充分」、「わからない」の場合は、「不十分」と「やや不十分」の選択率の合計を用いた。

データ公開経験と関連があると考えられる項目は多数(15項目)であったため、データを公開している研究者の特徴をよく説明する項目の組み合わせを明らかにするために、2項ロジスティック回帰分析を行った。目的変数はデータ公開経験の有無として、説明変数は関連がみられた項目から選択した。具体的な選択手順は結果の項で述べる(3.2.10)。

⁸⁾ Wordによる回答を入力の際に確認したところ、回答方法の誤り(択一式と択多式の誤りや、「公開経験あり」の回答者のみを対象としたフィルタークエスションの見落としなど)はなかった。

3.1.7 回答率と回答者の属性

調査依頼の送付数は1,983名、最終回答数は1,406名（70.9%）であった。このうち、3.1.4で言及した「研究を行ったことはない（口頭発表や論文出版の経験はない）」8名を除く1,398名（70.5%）の回答を本研究の分析対象とした。

回答者の年齢層と所属別の集計結果を表3-3に示す。所属は大学（60.3%）と企業（23.2%）の比率が高く、年齢層は31～40歳（46.3%）と41～50歳（29.4%）の比率がやや高かった。しかし、科学技術専門家ネットワークの構成比率と比較して、特に回答率が低い所属や年齢層はなかった。

表 3-3 回答者の年齢層と所属

年齢層	大学		企業		公的機関・団体		その他		合計	
	人数	比率	人数	比率	人数	比率	人数	比率	人数	比率
21-30 歳	11	0.8%	15	1.1%	4	0.3%	0	0.0%	30	2.1%
31-40 歳	411	29.4%	120	8.6%	114	8.2%	2	0.1%	647	46.3%
41-50 歳	249	17.8%	107	7.7%	54	3.9%	1	0.1%	411	29.4%
51-60 歳	103	7.4%	54	3.9%	31	2.2%	0	0.0%	188	13.4%
61-70 歳	69	4.9%	28	2.0%	20	1.4%	5	0.4%	122	8.7%
合計	843	60.3%	324	23.2%	223	16.0%	8	0.6%	1,398	100.0%

3 章 日本の研究者によるデータ公開状況と内的要因

回答者の分野と所属別の集計結果を表 3-4 に示す。排列は回答者数が多い分野の順とした。工学 (33.4%), 化学 (8.7%), 材料科学 (8.4%) の比率が高いが, 科学技術専門家ネットワークの構成比率と比較して, 特に回答率が低い分野はなかった⁹⁾。分野と年齢層別の集計結果は, 付表 46 に示す。

表 3-4 回答者の分野と所属

分野	大学		企業		公的機関・団体		その他		合計	
	人数	比率	人数	比率	人数	比率	人数	比率	人数	比率
工学	281	20.1%	136	9.7%	48	3.4%	2	0.1%	467	33.4%
化学	73	5.2%	41	2.9%	7	0.5%	0		121	8.7%
材料科学	62	4.4%	32	2.3%	24	1.7%	0		118	8.4%
生物学	52	3.7%	16	1.1%	17	1.2%	0		85	6.1%
複合領域	47	3.4%	10	0.7%	9	0.6%	0		66	4.7%
CS	45	3.2%	11	0.8%	6	0.4%	1	0.1%	63	4.5%
物理学	31	2.2%	8	0.6%	12	0.9%	0		51	3.6%
微生物学	22	1.6%	12	0.9%	14	1.0%	0		48	3.4%
神経科学	31	2.2%	8	0.6%	7	0.5%	0		46	3.3%
農業科学	24	1.7%	3	0.2%	19	1.4%	0		46	3.3%
環境学	23	1.6%	8	0.6%	13	0.9%	2	0.1%	46	3.3%
分子生物学	23	1.6%	9	0.6%	11	0.8%	1	0.1%	44	3.1%
地球科学	19	1.4%	5	0.4%	18	1.3%	0		42	3.0%
薬理学	23	1.6%	11	0.8%	1	0.1%	0		35	2.5%
臨床医学	18	1.3%	2	0.1%	2	0.1%	0		22	1.6%
植物動物学	13	0.9%	1	0.1%	5	0.4%	0		19	1.4%
数学	13	0.9%	2	0.1%	0		0		15	1.1%
免疫学	10	0.7%	1	0.1%	1	0.1%	0		12	0.9%
精神医学	9	0.6%	2	0.1%	2	0.1%	0		13	0.9%
社会科学	8	0.6%	1	0.1%	0		1	0.1%	10	0.7%
宇宙科学	3	0.2%	0		5	0.4%	0		8	0.6%
経済学	6	0.4%	0		1	0.1%	0		7	0.5%
その他	7	0.5%	5	0.4%	1	0.1%	1	0.1%	14	1.0%
合計	843	60.3%	324	23.2%	223	16.0%	8	0.6%	1,398	100.0%

⁹⁾ 最も回答率が高かったのは, 精神医学・心理学 (92.9%) であり, 最も低かったのは免疫学 (52.2%) であった。

3.2 結果

本節では、調査・分析結果を示す。Q35 の自由回答でオープンサイエンスに関する何らかの記述¹⁰⁾を行った回答者は 400 名であった。調査結果として、各質問の「その他」や「自由回答」(Q35) の記述もあわせて示す（以下、両者を区別しない場合は「自由記述」と記す）。

3.2.1 データの公開経験

研究データの公開経験と方法を確認するため、「これまでに、研究データを以下の方法で公開したご経験はありますか？」と尋ねた (Q18)。選択肢として「個人や研究室のウェブサイト」など、7 種類の公開方法を複数選択方式で示した。あわせて、7 種類の公開方法とは同時に選べない排他的選択肢¹¹⁾として「公開したことはない」、「わからない」、「研究にデジタルデータは用いない」を示した。表 3-5 の「あり」(713 名) は、7 種類の公開方法を 1 つ以上選択した回答者数の小計である。データの公開経験がある回答者は全体の 51.0%であった。「なし (公開したことはない)」は 46.9%、「わからない」は 1.9%、「研究にデジタルデータは用いない」は 0.1%であった。

表 3-5 データ公開経験の有無

データ公開経験	人数	比率
あり (公開方法を 1 つ以上選択)	713	51.0%
なし	656	46.9%
わからない	27	1.9%
研究にデジタルデータは用いない	2	0.1%
合計	1,398	100.0%

表 3-6 に、選択されたデータ公開方法を示す。公開方法は「個人や研究室のウェブサイト」(50.8%) や「論文の補足資料」(47.0%) の順に多く、「所属機関のリポジトリ」は 34.2%、「特定分野のリポジトリ」は 16.4%であった。「その他」には、“プロジェクト研究の Web サイト”などの記述があった。

¹⁰⁾ 謝辞などは除く。

¹¹⁾ ウェブアンケートシステムで設定したため、誤りによる無効回答などはなかった。3.1.5 で言及した Word ファイルによる提出分 (1 件) についても、回答方法の誤りはなかった。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-6 データを公開した方法 (n=713, 複数回答)

公開方法	人数	比率
個人や研究室のウェブサイト	362	50.8%
論文の補足資料 (supplementary materials)	335	47.0%
所属機関のリポジトリ・データアーカイブ	244	34.2%
特定分野のリポジトリ・データアーカイブ (DDBJ や ICPSR など)	117	16.4%
学術系 SNS (Mendeley, ResearchGate など)	70	9.8%
コード共有サービス (GitHub など)	24	3.4%
データ共有サービス (figshare, zenodo など)	16	2.2%
その他	18	2.5%

表 3-7 にデータ公開と OA のクロス集計結果を示す。OA が「あり」とする回答者は 991 名 (70.9%), 「なし」は 374 名 (24.8%), 「わからない」は 60 名 (4.3%) であった。つまり, OA の方がデータ公開よりも実践されていた。データ公開と OA 両方の経験がある回答者は 568 名 (40.6%), いずれの経験もない回答者は 214 名 (15.3%) であった。

表 3-7 データ公開と OA の経験の有無

データ \ OA	あり		なし		わからない		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
あり	568	40.6%	123	8.8%	22	1.6%	713	51.0%
なし	412	29.5%	214	15.3%	30	2.1%	656	46.9%
わからない	10	0.7%	9	0.6%	8	0.6%	27	1.9%
データは用いない	1	0.1%	1	0.1%			2	0.1%
合計	991	70.9%	347	24.8%	60	4.3%	1,398	100.0%

3.2.2 データの公開理由

データ公開や OA の経験がある回答者に, それぞれの理由を複数選択方式で尋ねた。結果を表 3-8 に示す。横線の下「あてはまるものはない」は, 排他的選択肢である (以降の表でも同様に示す)。

表 3-8 データ公開と OA の理由（複数回答）

	Q19. データ公開			Q3. OA		
	順位	人数	比率	順位	人数	比率
研究成果を広く認知してもらいたいから	1	417	58.5%	2	465	46.9%
論文を投稿した雑誌のポリシーだから [†]	2	312	43.8%	1	810	81.7%
科学研究や成果実装を推進したいから	3	186	26.1%	3	144	14.5%
他の研究者からのリクエストに応じて	4	181	25.4%	4	101	10.2%
所属機関のポリシーだから	5	169	23.7%	6	82	8.3%
オープンデータ/OA に貢献したいから	6	75	10.5%	5	94	9.5%
分野・コミュニティの規範だから	7	68	9.5%	7	29	2.9%
業績になる場合があるから [※]	8	66	9.3%	—	—	—
助成機関のポリシー（助成条件）だから	9	45	6.3%	8	20	2.0%
その他		14	2.0%		24	2.4%
あてはまるものはない		5	0.7%		1	0.1%
回答者数		713			991	

[†]論文を投稿した雑誌のポリシー（投稿規定）だから／論文を投稿した雑誌がオープンアクセスだから

[※]データのみの選択肢

データを公開した理由の1位は「研究成果を広く認知してもらいたいから」（58.5%）であった。表 3-9 に示す「その他」の自由記述でも、“次のポジションのための成果・技術アピール”といった記述がみられた。2位は「雑誌のポリシー」（43.8%）であり、OA では1位（81.7%）であった。科学的利他性に関連する「科学研究や成果実装の推進」は3位（26.1%）、「他の研究者からのリクエストに応じて」は4位（25.4%）、「オープンデータへの貢献」は6位（10.5%）であった。

7位の「分野・コミュニティの規範」（9.5%）は、OA の2.9%よりは選択率が高かったが、1割未満であった。分野別の内訳については3.2.8（分野による差の比較）で述べる。8位の「業績になる場合があるから」（9.3%）も1割に満たず、データ公開がほとんど評価に結びついていないことがわかった。国外で強い要因となっている「助成機関のポリシー」は9位（6.3%）であった。

表 3-9 に、「その他」の記述を「規範的」、「積極的・自発的」、「消極的・強制的」な理由ごとに示す。データについては、“それが当然のことと思っているから”という規範的な理由、研究の信頼性の担保（2名）、結果の検証を可能にする（2名）、新たな研究者とのつながり（1名）など、オープンサイエンスの目的と合致する積極的・自発的な理由、そして査読者・編集者に勧められた（2名）という理由がみられた。データと比較した論文の公開理由として特徴的で

3 章 日本の研究者によるデータ公開状況と内的要因

あったのは、積極的・自発的な理由であり、査読期間が短い（2名）、雑誌の IF、引用が増えるなど、業績や評価に直接関わるメリットがあると認識されていた。

表 3-9 データと論文の公開理由の自由記述

理由	Q19. データ	Q3. 論文
規範的	“それが当然のことと思っているから” (植物・動物学)	“知の共有による科学技術の推進加速 とムダの排除”
積極的 自発的	研究の信頼性の担保（2名）	査読期間が短い（2名）
	結果の検証を可能にする（2名）	先取権の確保
	“次のポジションのための成果・技術 アピール”	“OA の雑誌のインパクトファクターが 上昇”
	“データを公開すると実験設備の利用 料が安くなる”	“引用が増えると考えられたから”
消極的 強制的	“新たな研究者とのつながりの構築”	“多くの分野の学者や専門家と対話を 進める為”
	査読者・編集者に勧められた（2名）	OA 論文に選出された（2名）
		“知らない間に OA になっていた”

3.2.3 データの非公開理由と公開意思

データの公開経験および OA 経験が「なし」または「わからない」を選択した回答者に、その理由を複数選択方式で尋ねた（Q20）。結果を表 3-10 に示す。

表 3-10 データと論文の非公開理由（複数回答）

	Q20. データ			Q4. 論文		
	順位	人数	比率	順位	人数	比率
論文を投稿した雑誌のポリシーではないから [†]	1	180	26.4%	1	246	60.4%
時間が必要だから	2	177	25.9%	5	10	2.5%
所属機関にデータ公開/OA 方針がないから	3	153	22.4%	3	47	11.5%
ニーズがないと思うから	4	151	22.1%	6	9	2.2%
業績にならないから※	5	148	21.7%	—	—	—
リポジトリなどの公開手段がないから	6	61	8.9%	6	9	2.2%
資金が必要だから	7	52	7.6%	2	161	39.6%
助成機関のポリシー（助成条件）ではないから	8	42	6.1%	8	4	1.0%
分野・コミュニティで推奨されていないから	9	29	4.2%	4	13	3.2%
その他		61	8.9%		24	5.9%
あてはまるものはない		88	12.9%		28	6.9%
回答者数		683			407	

[†]論文を投稿した雑誌のポリシー（投稿条件）ではないから／投稿したい雑誌が OA ではないから

※データのみの選択肢

いずれも 1 位は「雑誌のポリシーではないから」であったが、データ公開は 26.4%、OA は 60.4%であり、比率が異なっていた。また、データ公開と OA の理由の順位はほぼ同一であったが、非公開理由の順位や選択率は異なっていた。データ公開は 1 位から 5 位までの選択率が 20%台であり、多くの回答者が選択した理由はなかった。一方、OA は 1 位の「雑誌のポリシーではないから」（60.4%）と 2 位の「資金が必要だから」（39.6%）¹²⁾に集中していた。

所属機関と助成機関のポリシーは、「所属機関にデータ公開方針がないから」（22.4%）が 3 位、「助成機関のポリシー（助成条件）ではないから」（6.1%）が 8 位であった。自由回答では、データ公開の義務化の潮流に対して、“研究者の選択を最大限に尊重すべき”など一律の義務化に抵抗を示す意見が 9 件みられた。

5 位の「業績にならないから」（21.7%）については、自由回答で“データ公開が業績評価の対象になるのであれば、多少は力を入れることも可能になります”など、所属機関や助成機関による業績化や評価（5 件）、引用（3 件）¹³⁾が保証されるならばデータ公開の可能性が高まることが示唆されていた。その一方で、“データを得ることは研究の一部ではあってもそれを

¹²⁾ 論文の投稿料や OA にするための APC の負担が大きいことは、自由回答でも多数指摘されていた。

¹³⁾ “データの公開は、誰がどういう目的でということがわかった上で、論文化時に引用等してもらえるなら歓迎します”など。

得ただけで業績として扱われるのは、問題であると思う”という意見もみられた。

「その他」には、(a)不要 (9 件)、(b)公開のインセンティブの欠如 (8 件)、(c)公開するデータがない (4 件) といった趣旨の記述がみられた。(a)データ公開は不要であるとする意見は、「論文や補足資料で充分である」、「必要ならば著者に直接問い合わせれば良い」、「必要性を感じられない」¹⁴⁾という 3 点に集約される。(b)公開のインセンティブの欠如について、自由回答においても同様の指摘が 18 件あり、“公開した人は公開の労力が必要なのに対してメリットが無く、そのデータを利用する側にメリットがあるため、公開する側に大きなメリットがある仕組み作りが必要”といった記述がみられた。インセンティブの具体例としては、先に述べた業績や経済的利益が示された。(c)公開するデータがないことについて、自由回答では分野に関連する指摘がみられた。これらは 3.2.8 (分野による差の比較) に示す。

続いて、非公開理由を 1 つ以上選択した回答者に「Q20 の理由が解決された場合、研究データを公開したいと思われますか？」と尋ねた (Q21)。また、OA についても同様に尋ねた (Q5)。表 3-11 に結果を示す。

表 3-11 データ公開と OA の意思

	Q21. データ公開		Q5. OA	
	人数	比率	人数	比率
はい	169	28.4%	296	78.1%
いいえ	130	21.8%	18	4.7%
わからない	296	49.7%	65	17.2%
回答者数	595	100.0%	379	100.0%

OA は「はい」の比率が 78.1%と高いが、データ公開は 28.4%と低く、「わからない」が約半数 (49.7%) を占めた。そこでデータ公開について「はい」、「いいえ」、「わからない」を選んだ回答者ごとに非公開理由を集計したが、偏りはみられなかった。つまり、データの非公開理由は雑誌や機関のポリシーがない、時間不足、ニーズがない、業績にならないなど多岐にわたり、かつ、解決されれば公開に繋がるような特定の理由がないことがわかった。

3.2.4 データ公開の障壁の程度

研究者からみたデータ公開の障壁の程度を明らかにするために、研究にデータを用いる回答者を対象として資源の充足状況 (Q28) や懸念 (Q29) について尋ね、データ公開経験の有

¹⁴⁾ 現時点で、研究者の立場からは再利用の可能性や保存の必要性を感じられないといった見解や、“データをパターン化されたシステムで公開する意義を感じない”といった記述がみられた。

無との関連を確認した。なお、Q28 と Q29 は尺度項目がそれぞれ 6 問ずつ続くため、回答の負担が大きいと考えられる。そこで、回答の信頼性を確認するために Cronbach のアルファ係数 α を算出したところ、Q28 は $\alpha=0.86$ 、Q29 は $\alpha=0.73$ であった。いずれの回答も、ある程度の内的一貫性を有していると考えられることから、これらの回答には一定の信頼性があるものと判断した。

[1] データを公開する場合の資源の状況

カレントデータを整備し、公開する場合の資源の充足状況を「不十分」、「やや不十分」、「ほぼ充分」、「充分」の 4 件法で尋ねた (Q28)。図 3-1 に、結果を「不十分」と「やや不十分」の選択率の合計が高い順に示す。なお、「データ保存用ストレージ」は、質問紙では「研究終了後のデータ保存用ストレージ」と記した ([付録 2] 参照)。

全体的に不十分であるという認識の回答者が多く、特にデータ整備・公開のための人材、資金、時間が不十分であるとする回答が多い。最も充足度が高い「研究中のデータ用ストレージ」についても、「充分」と「ほぼ充分」をあわせて 38.1% であった。また、データ公開用のリポジトリについては、「わからない」の比率が 27.1% と高く、「充分」と「ほぼ充分」をあわせても 18.4% にとどまった。

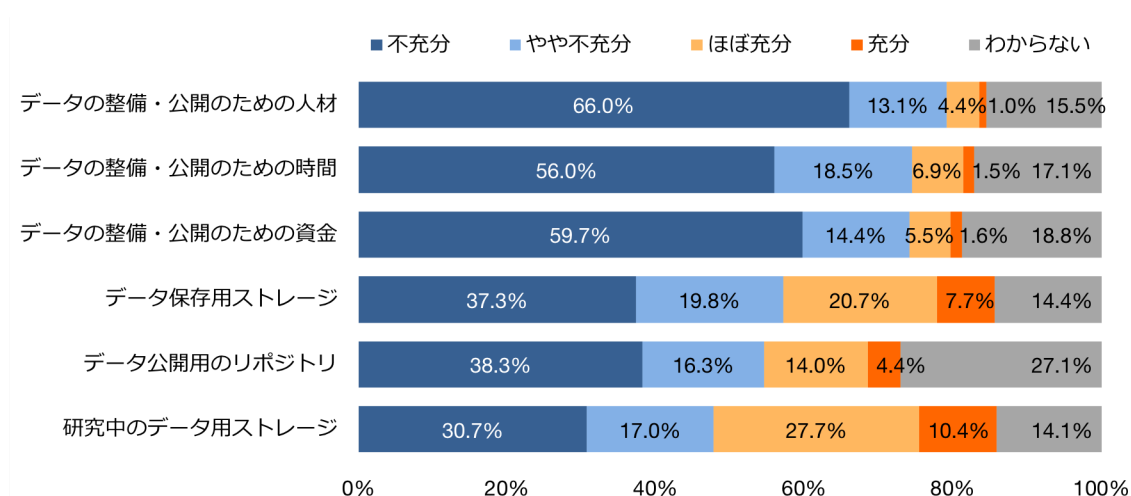


図 3-1 カレントデータの管理・公開資源の充足度 (n=1,396)

[2] データを公開する場合の懸念

カレントデータを公開する場合の懸念を「問題である」、「やや問題である」、「あまり問題ではない」、「問題ではない」の 4 件法で尋ねた (Q29)。図 3-2 に、結果を「問題である」と

3 章 日本の研究者によるデータ公開状況と内的要因

「やや問題である」の選択率の合計が高い順に示す。なお、「先に論文を出版される可能性」は、質問紙では「公開したデータを使って自分より先に論文を出版される可能性」と記した（〔付録 2〕参照）。

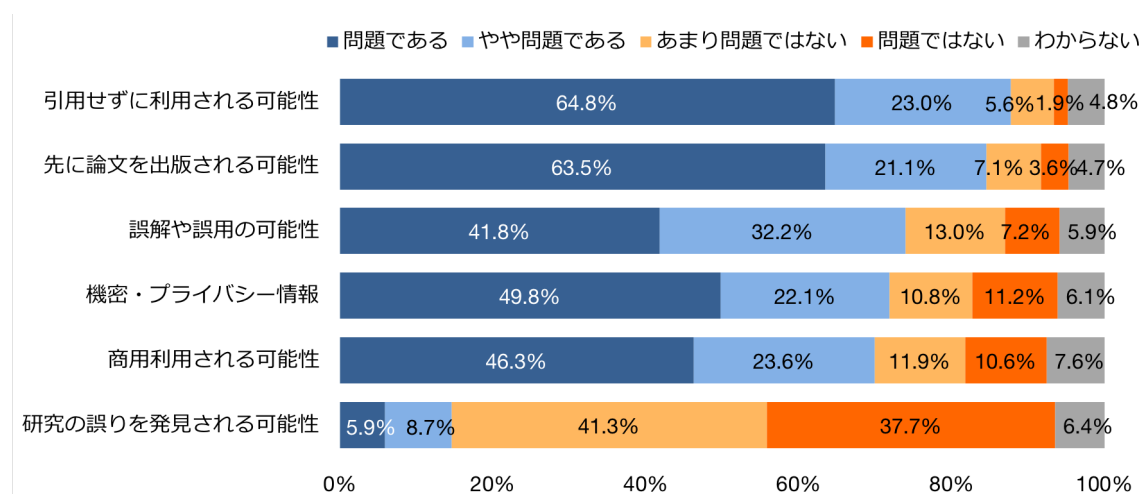


図 3-2 カレントデータを公開する場合の懸念（n=1,396）

全体的に「問題である」と認識している回答者が多いが、「研究の誤りを発見される可能性」だけは「問題である」と「やや問題である」をあわせても 14.6%であり、問題ではないと認識している回答者の比率が高かった。「引用せずに利用される可能性」は最も懸念が強かった。自由回答では、「盗用」、「剽窃」、「無断利用」、「フリーライダー」といった表現による懸念が示された。

論文の出版前の公開については、特許（17 件）、および先取権、業績評価、採用・昇進などのキャリアとの関連から否定的な見解が示され（10 件），“絶対にありえない”，“自殺行為である”など強い表現もみられた。論文出版後の公開であっても，“論文に掲載されたデータは非常に再公開しづらいため，（中略）厳しい就職・プロモーション環境の中ではマイナスの要因が強い”と抵抗を示す記述もみられた（2 件）。対策として，自由回答ではタイムスタンプなど公開日情報の管理（2 件）や“オリジナリティの担保”が必要であるとの記述がみられた。また，“特許に関連する情報については，「既発表」性に関する問題が広く認識され，どのような場合には公開を見送るべきか，など，研究者も適切に判断できるようになっていると思います”との見解も示されていた。

「誤解や誤用の可能性」については，単なる解釈の間違いではなく，“悪意のある引用”，“揚げ足取り”，“あらさがし”，“公開したことで血祭りにあげられるような事象”，“実験条件の差異により，公開データが間違っていると揶揄される”など利用者の態度や不適切な批判，および問い合わせへの対応コストを懸念する記述もみられた。対策として，利用規約の整備や，“利用する側の意識もきちんと教育される必要があります（中略）データの扱い，論文の

剽窃等、意識の面でグローバル・スタンダードに達しておらず、そのことが日本の学術に関する信用性を著しく低めていると感じています”などの意見がみられた。

3.2.5 データの提供と被提供経験

個人的なデータの提供状況がデータ公開経験と関連があるかどうかを確認するために、共同研究者を除く他の研究者へのデータの提供（Q9）および被提供状況（Q10）を尋ねた。表 3-12 にそれぞれの人数と比率を示す。データ提供の頻度の方が被提供の頻度よりもやや高かった。先行研究からデータを専ら提供する、あるいは提供される研究者がいるのではないかと想定していたが、回答を確認したところ、それぞれ 1 名ずつしか存在しなかった。また、データの提供・被提供状況の相関は高く（Spearman の相関係数 $r=0.771$, $p<0.001$ ），データを提供している回答者は提供も受けていることがわかった。

表 3-12 データの提供・被提供状況とデータ公開率

頻度	Q9. 提供		Q10. 被提供	
	人数	比率	人数	比率
よくある	104	7.4%	75	5.4%
たまにある	511	36.6%	466	33.4%
ほとんどない	469	33.6%	490	35.1%
まったくない	303	21.7%	359	25.7%
わからない	10	0.7%	7	0.5%
回答者数	1,397	100.0%	1,397	100.0%

3.2.6 公開データの入手経験と問題

[1] 公開データの入手経験

公開データを入手して再利用している研究者は、自身もデータを公開しているという先行研究（Piwowar, 2011）の結果を検証するために、入手経験と利用目的を尋ねた。まず、公開データの入手経験を確認するため、「これまでに、公開データを以下の公開先から入手したご経験はありますか？」と複数選択方式で尋ねた（Q11）。

表 3-13 にデータの入手先を 1 つ以上選択した回答者（以下、「データの入手経験がある回答者」）、および「ない」、「わからない」を選択した回答者の集計結果を示す。データの入手経験がある回答者は 1,060 名（75.8%）であり、回答者の約 3/4 は公開データの入手経験があ

3 章 日本の研究者によるデータ公開状況と内的要因

ることが明らかになった。

表 3-13 公開データの入手経験

データ入手経験	人数	比率
あり（1 つ以上を選択）	1,060	75.8%
なし	318	22.7%
わからない	20	1.4%
合計	1,398	100.0%

表 3-14 に公開データの入手先の結果を示す。選択率の順位は、表 3-6 に示したデータ公開先の順位と一致しており、個人や研究室のウェブサイト（64.8%）、論文の補足資料（53.1%）、学術機関のリポジトリ・データアーカイブ（50.0%）の順であった。特定分野のリポジトリ・データアーカイブの選択率は 1/4 に満たなかった（23.1%）。データ共有サービス（3.2%）やコード共有サービス（10.7%）、学術系 SNS（14.8%）の選択率は低かった。

表 3-14 公開データの入手方法（n=1,060, 複数回答）

入手方法	人数	比率
個人や研究室のウェブサイト	687	64.8%
論文の補足資料	563	53.1%
学術機関のリポジトリ・データアーカイブ（大学や NASA のリポジトリなど）	530	50.0%
特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）	245	23.1%
学術系 SNS（Mendeley, ResearchGate など）	157	14.8%
コード共有サービス（GitHub など）	113	10.7%
データ共有サービス（figshare, zenodo など）	34	3.2%
その他	13	1.2%

[2] 公開データの利用目的

入手した公開データの利用目的を「自身の研究のアイデアや仮説の参考にする（以下、「研究の参考」）」、「再分析・再利用して自身の研究を行う（以下、「研究に再利用」）」、「（他者の）研究を再現・追試する（以下、「研究を再現」）」に分けて、それぞれの頻度を尋ねた（Q13）。

図 3-3 に結果を示す¹⁵⁾。「行う」の比率、および「行う」と「たまに行う」の比率の合計は、いずれも「研究の参考」、「研究に再利用」、「研究を再現」の順に高かった。公開データの入手経験をもつ回答者のうち、「研究の参考」にした経験がある回答者の比率（「行う」と「たまに行う」の合計）は 91.3%、「研究に再利用」した経験がある回答者は 55.3%、「研究を再現」した経験がある回答者は 46.1%であった。

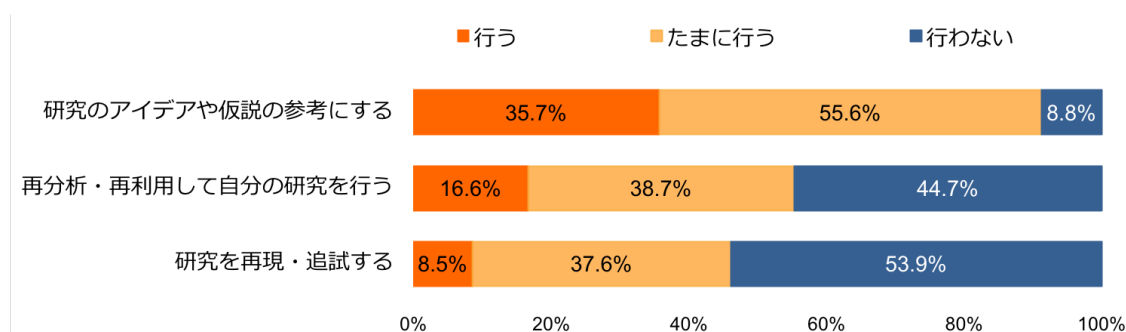


図 3-3 公開データの利用目的と頻度 (n=1,060)

[3] 公開データ入手における問題

公開データの入手における問題点を複数選択方式で尋ねた (Q12)。表 3-15 に結果を示す。問題を 1 つ以上選択した回答者は約 8 割 (79.8%) であった。

表 3-15 公開データ入手における問題の有無

データ入手の問題	人数	比率
あり (1 つ以上を選択)	846	79.8%
なし (問題を感じたことはない)	214	20.2%
合計	1,060	100.0%

表 3-16 に問題ごとの集計結果を示す。選択率が高かったのは「利用料金が必要」(43.1%)、次いで「利用者登録が必要」(33.3%)、「利用条件がよくわからない」(33.1%) であった。「その他」では、“データサイズが大きすぎてダウンロードできないことがある”など、データ量に関する問題 (3 名) や、“検索が一度にできずに、それぞれのデータベースに入る必要がある”など、統合検索システムの欠如 (2 名) についての記述がみられた。

¹⁵⁾ 比率を四捨五入したため、「研究のアイデアや仮説の参考にする」は合計が 100.1%になる。

表 3-16 公開データを入手する際の問題 (n=846, 複数回答)

公開データの入手の問題	人数	比率
利用料金が必要	365	43.1%
利用者登録が必要	282	33.3%
利用条件がよくわからない	280	33.1%
データごとに品質が異なる	272	32.2%
データごとにフォーマットが異なる	237	28.0%
利用したいデータへのアクセス方法がよくわからない	195	23.0%
著作者情報がよくわからない	172	20.3%
データの解釈や再利用方法がよくわからない	140	16.5%
リクエストから入手までに時間がかかる	139	16.4%
最新のデータを入手できない	120	14.2%
その他	18	2.1%

自由回答では、公開データの信頼性に対する指摘が多数みられた (12 件)。データ自体に査読が必要であるとする記述や、論文の出版前の公開は“データを利用する立場で考えても、論文化されていないデータはデータの質に不安があつて、利用したくない”、“review プロセスを経ていないので、データの信頼性に問題が残ると思われる”といった記述 (2 件) がみられた。

また、公開データの標準的なフォーマットが必要であるとする見解 (7 件) や、“統一フォーマットデータの提供を行えば、データのリサイクルやメタ解析分野の研究が進む”など活用を期待する記述もみられた。反面、標準フォーマットで公開するためのコストや“耐用年数”に疑問を呈する記述もみられた。このほか、国による不均衡¹⁶⁾ (2 件)、セキュリティ (2 件)、“失敗したデータ (と失敗した原因)”の公開といった課題が挙げられた。

[4] 公開データの検索ツールと情報源

公開データの探索方法を明らかにするために、公開データの入手経験がある回答者 1,034 名を対象として「公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい」と複数選択方式で尋ねた (Q15)。比較のために、論文についても全ての回答者 (1,398 名) を対象として同様に尋ねた (Q7)。ただし、データジャーナルは論文の探索方法に該当するものがないため、選択肢から除外した。

¹⁶⁾ 公開データを利用するだけで、データを公開しない国があるとの指摘がみられた。

表 3-17 にデータと論文の探索方法の集計結果を示す。これらの質問は回答を必須としておらず、無回答はデータが 1 名、論文が 4 名であった。

表 3-17 データと論文の探索方法（複数回答）

	Q15. データ			Q7. 論文		
	順位	人数	比率	順位	人数	比率
サーチエンジン（Google, Google Scholar など）	1	864	83.6%	1	1,164	83.5%
論文や学術記事の参考文献	2	734	71.1%	3	855	61.3%
出版社や学術雑誌のサイト（Elsevier, Wiley など）	3	391	37.9%	4	683	15.8%
研究者や同僚に尋ねる／教えてもらう	4	375	36.3%	6	469	33.6%
学術機関のリポジトリ・アーカイブ	5	338	32.7%	7	220	49.0%
データ／論文情報のデータベース [†]	6	187	18.1%	2	990	71.0%
学術系 SNS（Mendeley, ResearchGate など）	7	97	9.4%	8	134	9.6%
データジャーナル [※]	8	79	7.6%	—	—	—
政府・機関・出版社などの広報, ニュースレター	9	77	7.5%	10	98	7.0%
メーリングリスト	10	36	3.5%	12	64	4.6%
ブログや一般的な SNS（Facebook, Twitter など）	11	30	2.9%	11	72	5.2%
データ／論文のサーバ, リポジトリ [‡]	12	27	2.6%	5	675	48.4%
アラートサービス（RSS など）	13	19	1.8%	9	104	7.5%
その他		5	0.5%		11	0.8%
回答者数		1,033			1,394	

[†]データ情報のデータベース（Data Citation Index など）／論文情報のデータベース（Web of Science, Scopus, CiNii Articles など）

[‡]特定のデータリポジトリ／論文やプレプリントのサーバ（PubMed Central, arXiv, J-Stage など）

※データのみの選択肢（データ記述とデータへのリンクを掲載した雑誌）

データ、論文ともに最も使われているのはサーチエンジンであった。データの 2 位は「論文や学術記事の参考文献」（71.1%）であり、次いで「出版者や学術雑誌のサイト」（37.9%）であった。「データ／論文のデータベース（以下、「DB」）」と「データ／論文のサーバ, リポジトリ（以下、本項のみ「リポジトリ」と記す）」は、データと論文で順位と選択率が大きく異なった。DB はデータが 6 位（18.1%）であるのに対して論文は 2 位（71.0%）であり、リポジトリはデータが 12 位（2.6%）、論文は 5 位（48.4%）であった。

[5] 公開データの信頼性の判断基準

公開データの信頼性の判断基準としている事項を明らかにするために、公開データの入手経験がある回答者 1,034 名を対象として、複数選択方式で尋ねた (Q16)。比較のために、論文についても全ての回答者 (1,398 名) を対象として同様に尋ねた (Q8)。ただし、「そのデータを用いた論文」は除外した。

表 3-18 にデータと論文の信頼性の判断基準の集計結果を示す。これらの質問は回答を必須としておらず、無回答はデータが 9 名、論文が 4 名であった。

表 3-18 データと論文の信頼性の判断基準 (複数回答)

	Q16. データ			Q8. 論文		
	順位	人数	比率	順位	人数	比率
著者情報 (所属機関, 職位など)	1	733	71.5%	2	932	66.9%
研究手法の確かさ	2	649	63.3%	3	901	64.6%
そのデータを用いた論文※	3	602	58.7%	—	—	—
掲載リポジトリ/掲載雑誌	4	264	25.8%	1	1,219	87.4%
引用数 (そのデータを引用した文献数)	5	237	23.1%	4	479	34.4%
データにつけられた説明/抄録	6	221	21.6%	5	227	16.3%
ダウンロード数	7	54	5.3%	6	42	3.0%
オルトメトリクス	8	9	0.9%	7	5	0.4%
その他		20	2.0%		59	4.2%
回答者数		1,025			1,394	

※データのみの選択肢

データの信頼性の判断基準として、最も選択率が高いのは「著者情報」(71.5%) であった。一方、論文の信頼性の判断基準として、最も選択率が高いのは「掲載雑誌」(87.4%) であった。データを掲載しているリポジトリは、論文の掲載雑誌に相当すると仮定して選択肢に含めたが、これを選択した回答者は全体の 25.8% であった。また、「そのデータを用いた論文」は過半数 (58.7%) が選択していた。

「その他」には、「著者の過去論文, 過去の研究手法, 研究内容」など著者の評価に関する回答が 6 件、次いで“データそのものを解析してみる”, “データ間での比較による”などデータを直接確認するという回答が 4 件みられた。間接的な評価として, “研究者や同僚からのコメント”, “研究者コミュニティでの評価”があった。

[6] 利用したことがある公開データの分野と利用してみたい分野

公開データの利用状況や今後の利用意思を確認するため、S&EIに基づく13分野を示して複数選択方式で尋ねた。具体的には、利用経験がある回答者（1,060名）に「研究に利用したことがある（参考にしたり再分析した経験がある）公開データの分野をお選び下さい」と尋ね（Q14）、全ての回答者（1,398名）に「今後、利用してみたいと思う公開データの分野をお選び下さい」と尋ねた（Q17）。比較のため、全ての回答者に論文の利用状況も尋ねた（Q6）。

今後、利用してみたい公開データの分野について、1つ以上の分野を選択した回答者は97.1%であった。また、回答者一人あたりの平均選択数は2.79であり、データを利用したことがある分野（2.03）や論文（2.67）よりも有意に多かった。

表 3-19 に、利用してみたいと思う公開データの分野の選択率（%）を回答者の分野ごとに示す（行が回答者の分野、列が選択した分野）。なお、「その他」の分野は除外した。また、回答者自身の分野を選ぶ比率が高いため、他分野の回答者による選択率を、データの利用希望（「OD」）、論文の利用経験（「論文」）、データの利用経験（「DU」）のそれぞれについて示した。

表 3-19 公開データを利用してみたい分野（複数回答）（単位：％）

分野	n	工学	生物	化学	農学	医学	CS	物理	地球	社会	数学	心理	天文	人文	NU
工学	589	89.8	16.0	30.1	9.5	13.9	22.6	37.5	9.7	8.8	11.9	4.8	3.4	1.4	3.9
生物	180	12.2	92.2	18.9	41.7	55.6	28.3	11.1	10.0	3.9	5.6	6.7	1.1	2.2	1.7
化学	134	51.5	32.8	90.3	12.7	21.6	17.2	41.0	7.5	3.7	6.7	3.0	0.7	0.7	2.2
農学	123	18.7	78.9	30.1	90.2	36.6	14.6	4.9	16.3	9.8	3.3	2.4	0.8	1.6	0.8
医学	117	12.0	58.1	13.7	12.8	91.5	10.3	2.6	0.9	4.3	1.7	4.3	0.0	0.9	4.3
CS	71	47.9	14.1	5.6	7.0	19.7	90.1	11.3	1.4	16.9	18.3	15.5	0.0	11.3	1.4
物理	63	52.4	23.8	47.6	3.2	9.5	22.2	87.3	7.9	9.5	11.1	4.8	11.1	1.6	4.8
地球	55	36.4	16.4	21.8	25.5	5.5	20.0	27.3	92.7	14.5	12.7	1.8	7.3	7.3	0.0
社会	23	52.2	8.7	0.0	26.1	17.4	4.3	4.3	17.4	95.7	0.0	26.1	0.0	26.1	0.0
数学	18	38.9	0.0	5.6	5.6	5.6	55.6	33.3	5.6	16.7	66.7	0.0	11.1	5.6	11.1
心理	16	18.8	56.3	6.3	6.3	50.0	25.0	6.3	6.3	56.3	6.3	100	6.3	12.5	0.0
天文	4	0.0	0.0	25.0	0.0	0.0	25.0	75.0	50.0	0.0	0.0	0.0	100	0.0	0.0
人文	2	50.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0
合計	1,395	55.0	36.8	31.1	21.7	28.7	24.5	28.2	12.3	10.2	9.7	6.4	3.0	2.8	2.9
OD	1,395	29.5	28.6	24.8	15.1	22.9	21.0	25.5	9.0	8.7	8.9	5.3	2.7	2.7	
論文	1,395	32.9	27.3	26.9	13.1	20.6	14.9	27.6	7.1	6.1	8.8	3.9	2.1	1.5	
DU	1,057	22.0	21.9	17.9	7.8	12.6	10.4	16.7	4.5	2.7	3.4	1.2	1.4	1.2	

NU=研究にデータを使いたいとは思わない

OD=他分野の回答者による選択率

DU=公開データを使ったことがある分野のうち他分野の回答者による選択率

他分野からの選択率を分野別に比べると、13 分野全てについて、論文の利用率よりもデータの利用率が低かった。他分野からのデータ利用の希望率が高い分野は、工学（29.5%）、生物科学（28.6%）、物理学（25.5%）であった。また、論文の利用率と比較して、データ利用の希望率が高い分野、すなわち表 3-19 の「OD」と「論文」の差が大きかったのはコンピュータサイエンス（差は 6.1 ポイント）、社会科学（同 2.6 ポイント）、医学（同 2.3 ポイント）であった。

一方、論文の利用率と、データを利用してみたい分野の選択率を比べると、工学、化学、物理学を除く 10 分野はデータを利用してみたいとする回答の比率が高かった。この 3 分野は論文の利用率（表 3-19 の「論文」）とデータの利用率（同「DU」）の差が大きい分野でもあった（工学と物理学が 10.9 ポイント、化学が 9.0 ポイントで上位 3 分野を占める）。

3.2.7 データのプロファイル

研究者が扱うデータの種類, すなわち, (1)データ量, (2) データの所有権が複数の人や組織にあるかどうか, (3)データに機密情報が含まれるかどうか, (4)他の研究者によるデータの理解を難しいと考えているかどうかについて調査した。以下では, それぞれの結果について述べる。

[1] データの量

研究に用いるデータの量を明らかにするため, 「カレントデータの総量は, およそどれくらいでしたか? 論文などには使用しなかったデータも含めてあてはまる単位をお選びください。」と尋ねた (Q23)。図 3-4 に示す通り, GB (ギガバイト) が最も多く 56.1%, 次いで MB (メガバイト) が 18.8%, TB (テラバイト) が 13.5%であった。PB (ペタバイト) 以上を選択した回答者は, 0.3%であった。この質問は回答を必須としておらず, 無回答は 3 名であった。

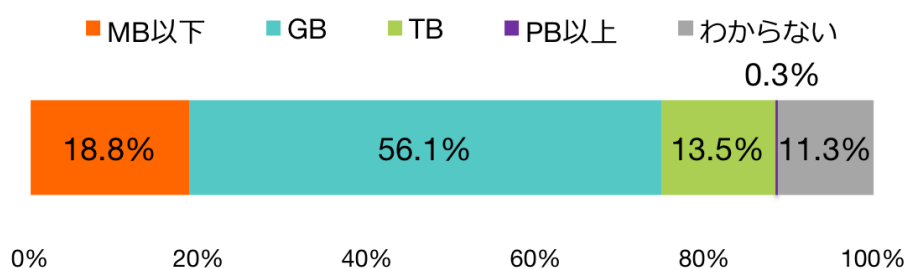


図 3-4 カレントデータの量 (n=1,393)

[2] データの所有権

カレントデータの所有権が回答者以外にあるかどうかと, その種類を複数選択方式で尋ねた (Q24)。表 3-20 に結果を示す。この質問は回答を必須としておらず, 無回答は 1 名であった。回答者の 8 割以上 (83.6%) が所有権をもつ人・組織を 1 つ以上選択しており, 「なし (自分のみ)」は 8.2%, 「わからない」は 8.2%であった。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-20 カレントデータの所有権をもつ人・組織の有無

所有権をもつ人・組織	人数	比率
あり（1 つ以上を選択）	1,166	83.6%
なし（自分のみ）	115	8.2%
わからない	114	8.2%
合計	1,395	100.0%

表 3-21 に選択肢ごとの集計結果を示す。最も多いのが共同研究者（82.1%）、次いで所属機関（58.2%）であった。「その他」には、“データ提供した実験参加者個々人”，“公開されているデータベースの権利者”，“研究プロジェクトの発注機関”，“出版社に譲渡している中に含まれているものと考えていた”といった記述がみられた。

表 3-21 カレントデータの所有権をもつ人・組織（n=1,166, 複数回答）

所有権をもつ人・組織	人数	比率
共同研究者	957	82.1%
所属機関	679	58.2%
所属機関以外の研究機関	111	9.5%
所属機関以外の企業	75	6.4%
研究助成機関	67	5.7%
その他	7	0.6%

[3] データの機密情報

データを公開する場合の懸念の 1 つである機密情報が、カレントデータに含まれているかどうかとその種類を尋ねた（Q25）。結果を表 3-22 に示す。この質問は回答を必須としておらず、無回答は 13 名であった。回答者の約半数（49.4%）が機密情報を 1 つ以上選択しており、「なし」は 35.6%、「わからない」は 15.0%であった。

表 3-22 カレントデータの機密情報の有無

機密情報	人数	比率
あり（1 つ以上を選択）	683	49.4%
なし	493	35.6%
わからない	207	15.0%
合計	1,396	100.0%

表 3-23 に選択肢ごとの集計結果を示す。選択肢には、先行研究などで機密情報とされている個人情報や商業上の機密情報などを挙げたが、「その他、守秘義務がある情報」が最も多かった（58.4%）。なお、機密に関わるため、この質問のみ「その他」で具体的な記述を求めなかった。選択肢では、個人情報（42.9%）、企業・商業上の機密情報（42.2%）の順に多かった。

表 3-23 カレントデータに含まれる機密情報（n=683、複数回答）

機密情報	人数	比率
個人情報	293	42.9%
企業・商業上の機密情報	288	42.2%
健康情報（遺伝情報、医療情報）	76	11.1%
その他、守秘義務がある情報	399	58.4%

[4] データの理解の難しさ

カレントデータを公開した場合に、同じ分野の研究者や理解できるかどうかを 3 件法で尋ねた（Q30）。同様に、異分野の研究者が理解できるかどうかも尋ねた（Q31）。

図 3-31 に結果を示す。データを公開した場合、同じ分野の研究者は理解できると考える回答者が多く、異分野の研究者は難しいと考える回答者が多かった。同じ分野の研究者は 60.0% が「できると思う」を選択していたのに対して、異分野の研究者については 13.2% にとどまった。一方、同じ分野の研究者は「難しいと思う」を選択していたのが 10.4% であったのに対して、異分野の研究者については 42.8% が選択していた。

3 章 日本の研究者によるデータ公開状況と内的要因

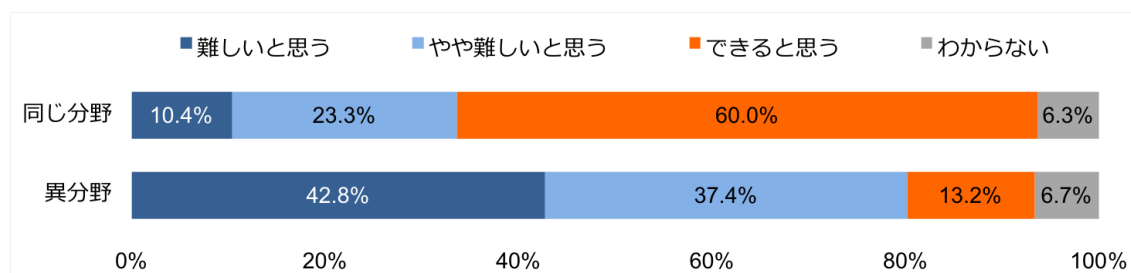


図 3-5 カレントデータの理解：同じ分野と異分野 (n=1,396)

3.2.8 分野による差の比較

3.2.1 から 3.2.7 までの項目について、分野ごとに集計した。以下では差がみられた項目を、それぞれ示す。また、次節（3.3 考察）の表 3-61 に、差がみられた項目の一覧を示す。

[1] 分野別データ公開経験

図 3-6 に、分野別のデータ公開経験を示す。データ公開経験がある回答者の比率が高い分野は、植物・動物学 (84.2%)、分子生物学・遺伝学 (70.5%)、環境・生態学 (69.6%) であり、低い分野は精神医学・心理学 (30.8%)、工学 (41.1%)、経済学・経営学 (42.9%) であった。

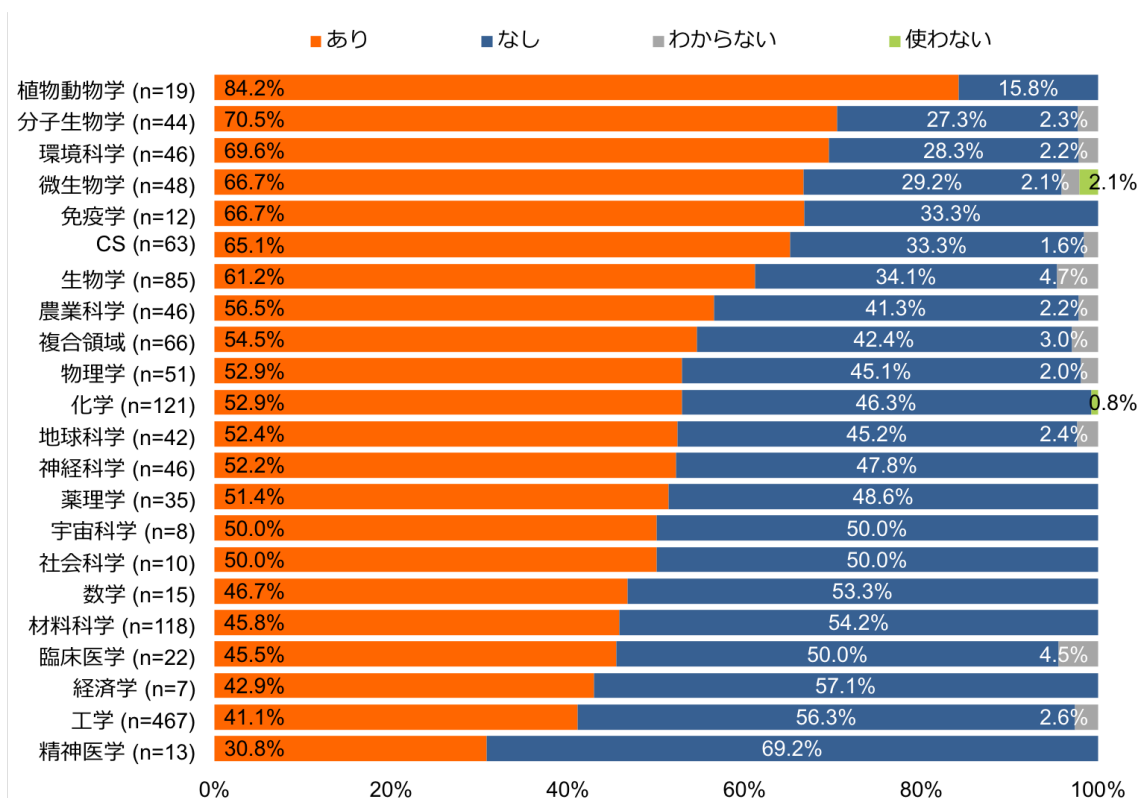


図 3-6 分野別データ公開経験 (n=1,384)

表 3-24 に分野別の回答者数とデータ公開率、および関連する主な結果を示す。排列は、データ公開経験「あり」の比率が高い順とした。なお、分野が「その他」の回答者を除去しているため、全体の結果と本項の結果は回答者数 (n) が異なる。

「データ」はデータ公開経験、「OA」は OA 論文の経験、「意思」はデータの非公開理由が解決された場合の公開意思、「入手」は公開データの入手経験、「機密」はカレントデータに機密やプライバシー情報が、それぞれ「あり」の回答者の比率を示している。「提供」は、データの提供経験が「よくある」と「たまにある」の回答者の比率の合計を示している。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-24 分野別データ公開率および関連する結果

分野	n	データ	OA	意思	提供	入手	機密
植物動物学	19	84.2%	84.2%	33.3%	57.9%	89.5%	5.3%
分子生物学	44	70.5%	77.3%	30.0%	52.3%	97.7%	43.2%
環境学	46	69.6%	84.8%	25.0%	60.0%	73.9%	37.8%
微生物学	48	66.7%	72.9%	15.4%	41.7%	85.4%	50.0%
免疫学	12	66.7%	75.0%	0.0%	35.5%	75.0%	75.0%
CS	63	65.1%	74.6%	35.3%	52.4%	85.7%	47.5%
生物学	85	61.2%	70.6%	19.4%	38.8%	81.2%	45.9%
農業科学	46	56.5%	78.3%	15.8%	50.0%	80.4%	32.6%
複合領域	66	54.5%	77.3%	14.8%	50.0%	75.8%	47.0%
化学	121	52.9%	57.0%	28.3%	30.6%	76.9%	48.3%
物理学	51	52.9%	88.2%	27.3%	43.2%	74.5%	47.1%
地球科学	42	52.4%	85.7%	47.1%	66.6%	90.5%	36.6%
神経科学	46	52.2%	78.3%	35.0%	39.1%	82.6%	34.8%
薬理学	35	51.4%	80.0%	13.3%	20.0%	71.4%	58.8%
社会科学	10	50.0%	80.0%	0.0%	40.0%	80.0%	30.0%
宇宙科学	8	50.0%	75.0%	66.7%	87.5%	87.5%	25.0%
数学	15	46.7%	73.3%	0.0%	66.7%	73.3%	20.0%
材料科学	118	45.8%	69.5%	33.9%	46.6%	77.1%	60.2%
臨床医学	22	45.5%	72.7%	33.3%	36.4%	50.0%	68.2%
経済学	7	42.9%	85.7%	0.0%	42.9%	71.4%	42.9%
工学	467	41.1%	63.8%	32.2%	43.7%	70.0%	55.0%
精神医学	13	30.8%	100.0%	33.3%	23.1%	46.2%	69.2%
全体		51.2%	70.9%	28.7%	44.2%	76.0%	49.3%
回答者数※	1,384	1,384	1,384	586	1,383	1,384	1,369

「データ」データ公開経験, 「OA」OA 経験, 「意思」データの非公開理由が解決した場合に公開する, 「提供」データの提供経験, 「入手」公開データの入手経験, 「機密」カレントデータに機密やプライバシー情報が含まれる回答者の比率

※分野で「その他」を選択した回答者を除く。

表 3-7 に示したように, データ公開と OA 両方の経験がある回答者は 40.6%, いずれもない回答者は 15.3%であった。しかし, 分野別のデータ公開経験がある回答者の比率と OA 経験がある回答者の比率 (表 3-24) について Pearson の相関係数を用いて確認したが, 相関は

認められなかった。たとえば、精神医学・心理学はデータ公開経験をもつ回答者は 30.8% (平均 51.2%) であったが、OA は 100.0% (平均 70.9%) であった。つまり、OA 経験をもつ回答者が多い分野はデータ公開経験をもつ回答者も多いとはいえなかった。

[2] データの公開理由

表 3-25 に、分野別にデータ公開経験がある回答者数 (n) と各理由の選択率 (回答者数の合計 708 名) を示す。排列は、表 3-24 に示したデータ公開率が高い順とした。表では、「研究成果を広く認知してもらいたいから」を「認知」、「論文を投稿した雑誌のポリシー (投稿規定) だから」を「雑誌」、「科学研究や成果実装を推進したいから」を「実装」、「他の研究者からのリクエストに応じて」を「リク」、「所属機関のポリシーだから」を「所属」、「オープンデータに貢献したいから」を「貢献」、「分野・コミュニティの規範だから」を「規範」、「業績になる場合があるから」を「業績」、「助成機関のポリシー (助成条件) だから」を「助成」と記す。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-25 分野別データの公開理由

分野	n	認知	雑誌	実装	リク	所属	貢献	規範	業績	助成
植物動物学	16	75.0%	62.5%	43.8%	37.5%	6.3%	37.5%	25.0%	18.8%	6.3%
分子生物学	31	48.4%	80.6%	22.6%	22.6%	12.9%	19.4%	22.6%	9.7%	12.9%
環境学	32	43.8%	62.5%	43.8%	34.4%	15.6%	21.9%	15.6%	12.5%	6.3%
免疫学	8	50.0%	62.5%	37.5%	12.5%	12.5%	25.0%	12.5%	0.0%	25.0%
微生物学	32	56.3%	56.3%	15.6%	6.3%	15.6%	6.3%	12.5%	12.5%	9.4%
CS	41	82.9%	22.0%	46.3%	34.1%	17.1%	22.0%	17.1%	14.6%	7.3%
生物学	52	59.6%	67.3%	17.3%	19.2%	15.4%	9.6%	5.8%	1.9%	7.7%
農業科学	26	50.0%	38.5%	11.5%	19.2%	38.5%	3.8%	7.7%	3.8%	7.7%
複合領域	36	63.9%	44.4%	30.6%	19.4%	25.0%	16.7%	13.9%	13.9%	5.6%
化学	64	45.3%	42.2%	15.6%	21.9%	28.1%	7.8%	7.8%	10.9%	3.1%
物理学	27	85.2%	37.0%	51.9%	25.9%	18.5%	14.8%	3.7%	3.7%	3.7%
地球科学	22	68.2%	22.7%	40.9%	45.5%	22.7%	13.6%	18.2%	18.2%	4.5%
神経科学	24	41.7%	54.2%	20.8%	12.5%	20.8%	8.3%	16.7%	12.5%	0.0%
薬理学	18	66.7%	66.7%	11.1%	5.6%	33.3%	5.6%	0.0%	5.6%	0.0%
社会科学	5	60.0%	40.0%	40.0%	20.0%	80.0%	0.0%	0.0%	20.0%	0.0%
宇宙科学	4	100.0%	25.0%	100.0%	75.0%	0.0%	25.0%	0.0%	0.0%	0.0%
数学	7	28.6%	28.6%	14.3%	42.9%	14.3%	0.0%	28.6%	28.6%	0.0%
材料科学	54	59.3%	53.7%	24.1%	27.8%	22.2%	5.6%	7.4%	1.9%	5.6%
臨床医学	10	50.0%	40.0%	20.0%	30.0%	20.0%	0.0%	10.0%	30.0%	10.0%
経済学	3	66.7%	33.3%	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%
工学	192	57.8%	28.1%	23.4%	28.1%	31.3%	6.3%	4.2%	7.8%	7.3%
精神医学	4	50.0%	50.0%	0.0%	25.0%	25.0%	0.0%	0.0%	0.0%	0.0%
全体	708	58.5%	43.8%	26.1%	25.3%	23.9%	10.6%	9.5%	9.2%	6.4%

認知＝研究成果を広く認知してもらいたいから，雑誌＝雑誌のポリシーだから，実装＝科学研究や成果実装を推進したいから，リク＝他の研究者からのリクエストに応じて，所属＝所属機関のポリシーだから，貢献＝オープンデータに貢献したいから，規範＝分野・コミュニティの規範だから，業績＝業績になる場合があるから，助成＝助成機関のポリシーだから

全体で最も選択率が高かった「研究成果を広く認知してもらいたいから（以下，「認知）」（全体 58.5%）は，宇宙科学（100.0%），物理学（85.2%），コンピュータサイエンス（82.9%），植物・動物学（75.0%）の順に高く，数学（28.6%），神経科学（41.7%），環境・生態学（43.8%），化学（45.3%）は低かった。分野別のデータ公開率と「認知」の選択率に関連があるかどうか

を Pearson の相関係数を用いて確認したが、相関は認められなかった。

科学的利他性に関連する「科学研究や成果実装を推進したいから」、「他の研究者からのリクエストに応じて」、「オープンデータに貢献したいから」について述べる。「科学研究や成果実装を推進したいから」（全体 26.1%）は、宇宙科学（100.0%）、物理学（51.9%）、コンピュータサイエンス（46.3%）、植物・動物学および環境・生態学（43.8%）の順に高く、上位 4 分野は「研究成果を広く認知してもらいたいから」と同様であった。一方、経済学・経営学と精神医学・心理学では全く選択されていなかった。「他の研究者からのリクエストに応じて」（全体 25.3%）は、宇宙科学（75.5%）、地球科学（45.5%）、数学（42.9%）であった。薬理学・毒性学（5.6%）、微生物学（6.3%）、神経科学・行動学（12.5%）であり、全く選択されていない分野はみられなかった。つまり、全ての分野で他の研究者からのリクエストに応じてデータを公開した経験をもつ回答者がいることがわかった。「オープンデータに貢献したいから」（全体 10.6%）は、植物・動物学（37.5%）、免疫学（25.0%）、宇宙科学（25.0%）、コンピュータサイエンス（22.0%）の順に高く、臨床医学、精神医学・心理学、経済学・経営学、数学、社会科学の 5 分野では全く選択されていなかった¹⁷⁾。

続いてデータ公開を求める制度である雑誌、所属機関、助成機関のポリシーについて述べる。雑誌のポリシーは、分子生物学・遺伝学（80.6%）、生物学・生化学（67.3%）、薬理学・毒性学（66.7%）の順に選択率が高く、コンピュータサイエンス（22.0%）、地球科学（22.7%）、宇宙科学（25.0%）は低かった。データ公開経験とは中程度の相関が認められた（Pearson の相関係数 $r=0.488, p<0.01$ ）。つまり、データ公開率が高い分野の回答者は、その理由として「雑誌のポリシーだから」を選択する傾向がみられた。所属機関のポリシーは、全体的に選択率が低かった（23.9%）が、社会科学（80.0%）は突出して高く、農業科学（38.5%）、薬理学・毒性学（33.3%）と差がみられた。一方、経済学・経営学と宇宙科学は全く選択されていなかった。データ公開率との相関は認められなかった。助成機関のポリシーも選択率が低く（全体は 6.4%）、特に選択率が高い分野もみられなかった。選択率が高い順に、免疫学（25.0%）、分子生物学・遺伝学（12.9%）、臨床医学（10.0%）であり、経済学・経営学、数学など、7 分野は全く選択されていなかった¹⁷⁾。

全体 7 位の「分野・コミュニティの規範だから（以下、「規範」）」（全体 9.5%）は、選択率が高い分野であっても 3 割未満（数学 28.6%、植物・動物学 25.0%、分子生物学・遺伝学 22.6%）であった。図 3-7 に示すように、「規範」の選択率と、全体 8 位の「業績になる場合があるから（以下、「業績」）」（9.2%）の選択率の関連を確認したところ、規範であると認識されている分野では、おおむね業績として評価されると認識されている傾向がみられた（Pearson の相関係数は $r=0.596, p<0.01$ ）。ただし、臨床医学は「規範」が 10.0%で「業績」が 30.0%と差が

¹⁷⁾ 「オープンデータに貢献したいから」、「分野・コミュニティの規範だから」、「助成機関のポリシーだから」はデータ公開率と選択率には有意な相関がみられたが、選択率が低く、全く選択していない分野も複数みられたため、関連があるとはいえないと考えた。

3 章 日本の研究者によるデータ公開状況と内的要因

大きく、社会科学は「規範」が0%で「業績」が20.0%，薬理学・毒性学は「規範」が0%で「業績」が5.6%であり，逆に免疫学は「規範」が12.5%で「業績」が0%であった。精神医学・心理学，経済学・経営学，宇宙科学は，「規範」と「業績」ともに選択率が0%であった。

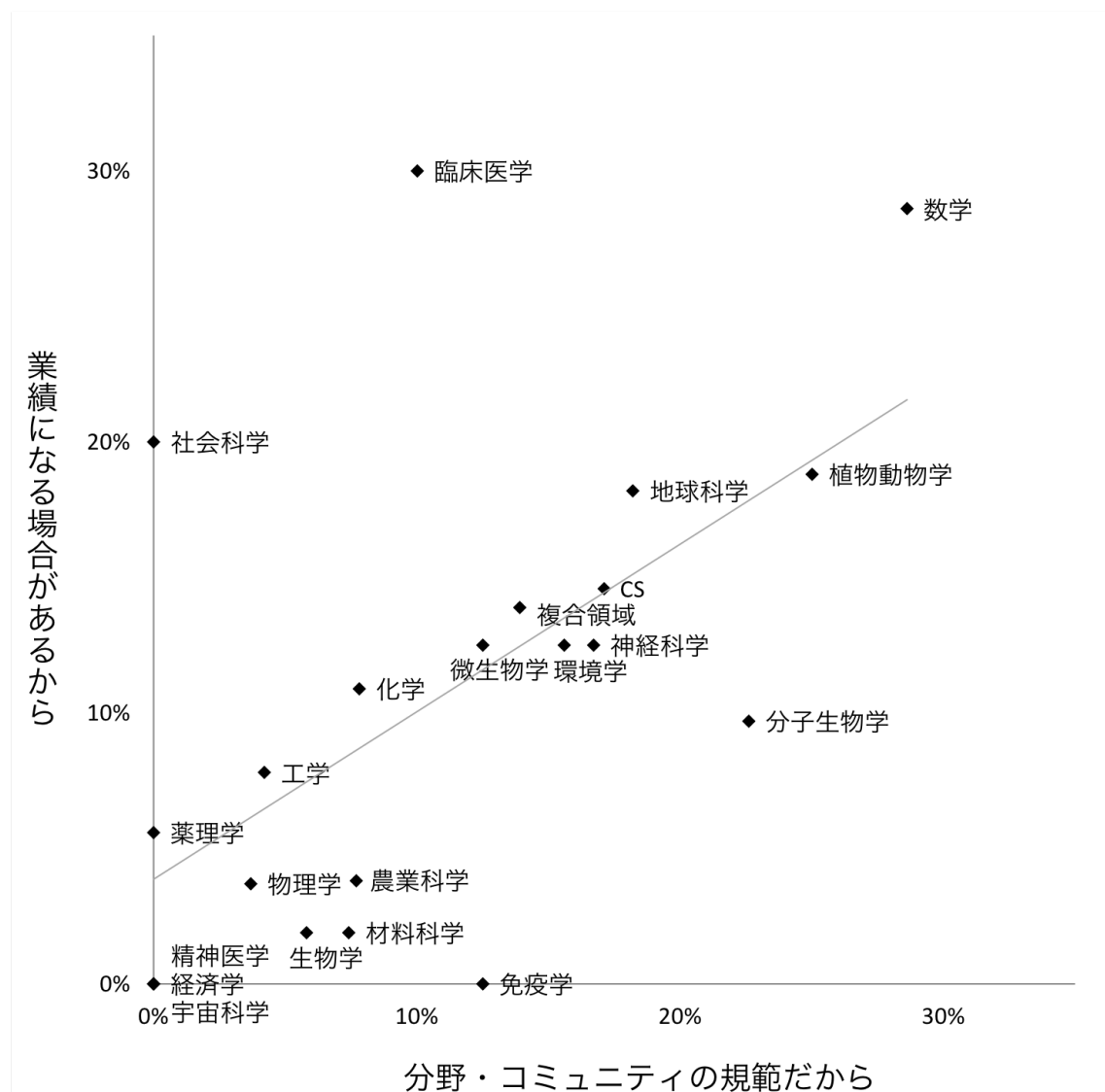


図 3-7 分野別データ公開の誘因（規範と業績）

[3] データの非公開理由と公開意思

表 3-26 に，分野別のデータ公開経験がない回答者数（n）と非公開理由の選択率を示す。排列は，表 3-24 に示したデータ公開率が低い順とした。表では，「論文の投稿した雑誌のポリシー（投稿条件）ではないから」を「雑誌」，「時間が必要だから」を「時間」，「所属機関

にデータ公開ポリシーがないから」を「所属」,「ニーズがないと思うから」を「ニーズ」,「業績にならないから」を「業績」,「リポジトリなどの公開手段がないから」を「手段」,「資金がないから」を「資金」,「助成機関のポリシー（助成条件）ではないから」を「助成」,「分野・コミュニティで推奨されていないから」を「分野」と記す。

表 3-26 分野別データの非公開理由

分野	n	雑誌	時間	所属	ニーズ	業績	手段	資金	助成	分野
精神医学	9	11.1%	44.4%	22.2%	22.2%	22.2%	44.4%	11.1%	0.0%	0.0%
工学	275	25.5%	25.5%	25.5%	18.5%	17.5%	6.5%	8.0%	3.6%	3.3%
経済学	4	50.0%	75.0%	25.0%	50.0%	50.0%	25.0%	50.0%	50.0%	25.0%
臨床医学	12	16.7%	33.3%	0.0%	16.7%	8.3%	8.3%	16.7%	8.3%	8.3%
材料科学	64	28.1%	28.1%	26.6%	26.6%	25.0%	7.8%	9.4%	4.7%	3.1%
数学	8	0.0%	0.0%	0.0%	25.0%	12.5%	0.0%	0.0%	0.0%	25.0%
社会科学	5	20.0%	20.0%	20.0%	40.0%	0.0%	0.0%	0.0%	0.0%	20.0%
宇宙科学	4	25.0%	25.0%	25.0%	25.0%	0.0%	0.0%	25.0%	0.0%	0.0%
薬理学	17	41.2%	17.6%	29.4%	23.5%	23.5%	5.9%	5.9%	11.8%	5.9%
神経科学	22	31.8%	27.3%	4.5%	45.5%	27.3%	9.1%	4.5%	4.5%	4.5%
地球科学	20	20.0%	50.0%	20.0%	25.0%	15.0%	10.0%	5.0%	15.0%	0.0%
化学	56	21.4%	16.1%	21.4%	23.2%	21.4%	7.1%	3.6%	3.6%	5.4%
物理学	24	33.3%	16.7%	37.5%	20.8%	29.2%	12.5%	0.0%	8.3%	4.2%
複合領域	30	20.0%	46.7%	26.7%	23.3%	43.3%	3.3%	16.7%	13.3%	3.3%
農業科学	20	45.0%	15.0%	15.0%	15.0%	35.0%	25.0%	5.0%	5.0%	5.0%
生物学	33	42.4%	24.2%	9.1%	24.2%	30.3%	12.1%	12.1%	6.1%	3.0%
CS	22	31.8%	40.9%	9.1%	9.1%	13.6%	18.2%	13.6%	4.5%	4.5%
免疫学	4	25.0%	0.0%	25.0%	0.0%	50.0%	0.0%	0.0%	25.0%	25.0%
微生物学	15	20.0%	6.7%	46.7%	33.3%	13.3%	6.7%	0.0%	33.3%	0.0%
環境学	14	14.3%	35.7%	14.3%	7.1%	21.4%	14.3%	0.0%	7.1%	7.1%
分子生物学	13	7.7%	23.1%	15.4%	46.2%	23.1%	7.7%	0.0%	0.0%	0.0%
植物動物学	3	33.3%	33.3%	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	33.3%
全体	674	26.3%	26.3%	22.4%	22.1%	21.5%	8.8%	7.7%	6.1%	4.3%

雑誌＝雑誌のポリシーではないから, 時間＝時間が必要だから, 所属＝所属機関にデータ公開ポリシーがないから, ニーズ＝ニーズがないと思うから, 業績＝業績にならないから, 手段＝リポジトリなどの公開手段がないから, 資金＝資金がないから, 助成＝助成機関のポリシーではないから 分野＝分野・コミュニティで推奨されていないから

3 章 日本の研究者によるデータ公開状況と内的要因

分野別のデータの非公開理由とデータ公開率の関連を調べたが、いずれの理由についても有意な相関はみられなかった。つまり、データ公開率が低い分野の回答者による選択率が高い、あるいは低い非公開理由はなかった。たとえば、「雑誌のポリシーではないこと」は経済学・経営学 (50.0%)、農業科学 (45.0%)、生物学・生化学 (42.4%) の順に選択率が高く、低かったのは、数学 (0.0%)、分子生物学・遺伝学 (7.7%)、精神医学・心理学 (11.1%) の順であった。相対的にデータ公開率が低い数学 (46.7%) は全く選択しておらず、公開率が高い生物学・生化学 (61.2%) の選択率が高かった。所属機関のポリシーではないことは、微生物学 (46.7%)、物理学 (37.5%)、薬理学・毒性学 (29.4%) の順に選択率が高く、臨床医学、数学、植物・動物学では全く選択されていなかった。助成機関のポリシーではないことは、経済学・経営学 (50.0%)、微生物学 (33.3%)、免疫学 (25.0%) の順に選択率が高く、数学、分子生物学・遺伝学、植物・動物学、精神医学・心理学、社会科学、宇宙科学では全く選択されていなかった。

自由回答では、“大掛かりな実験をしなければ再現できない実験系の分野と、読者が自分の頭で考えて再現できる理論系の分野では事情が全く異なると思われる”、“数学は純粋に理論的な研究分野であるため、実証研究的な意味合いでのデータというものは存在せず、(数表・分類表などのように) 特定の数学的事実を網羅的にまとめたものや、抽象的・一般的事実の具体例を提示するものとしてのみデータが扱われることになる”、“(物性物理学は) 生データを公開しても他研究者に理解することは難しい。理解できるように整えるのは、時間的にかなり負担である。また、結果を分かりやすくまとめたものが論文であり、生データを公開する意味が分からない (この点は、分野によって事情が異なるのかもしれませんが)”など、分野によってはデータ公開の必要性やコストパフォーマンスが低いことについての記述がみられた。また、工学分野は製品やノウハウなど企業機密に近いため困難であるという記述がみられた (5 件)。

表 3-27 に、分野別のデータ公開意思を示す。排列は「はい」の比率が高い順として、同順位の場合は「いいえ」の比率が低い順とした。

表 3-27 分野別データ公開意思

分野	回答者数	はい		いいえ		わからない	
宇宙科学	3	2	66.7%	1	33.3%	0	0.0%
地球科学	17	8	47.1%	1	5.9%	8	47.1%
CS	17	6	35.3%	2	11.8%	9	52.9%
神経科学	20	7	35.0%	7	35.0%	6	30.0%
材料科学	59	20	33.9%	13	22.0%	26	44.1%
臨床医学	9	3	33.3%	1	11.1%	5	55.6%
精神医学・心理学	9	3	33.3%	1	11.1%	5	55.6%
植物・動物学	3	1	33.3%	1	33.3%	1	33.3%
工学	236	76	32.2%	48	20.3%	112	47.5%
分子生物学・遺伝学	10	3	30.0%	2	20.0%	5	50.0%
化学	46	13	28.3%	14	30.4%	19	41.3%
物理学	22	6	27.3%	7	31.8%	9	40.9%
環境・生態学	12	3	25.0%	0	0.0%	9	75.0%
生物学・生化学	31	6	19.4%	5	16.1%	20	64.5%
農業科学	19	3	15.8%	5	26.3%	11	57.9%
微生物学	13	2	15.4%	4	30.8%	7	53.8%
複合領域	27	4	14.8%	9	33.3%	14	51.9%
薬理学	15	2	13.3%	2	13.3%	11	73.3%
経済学・経営学	4	0	0.0%	0	0.0%	4	100.0%
免疫学	3	0	0.0%	0	0.0%	3	100.0%
社会科学	5	0	0.0%	1	20.0%	4	80.0%
数学	6	0	0.0%	3	50.0%	3	50.0%
合計	586	168	28.7%	127	21.7%	291	49.7%

非公開理由が解決された場合に公開したいと考える回答者の比率は、宇宙科学（66.7%）、地球科学（47.1%）、コンピュータサイエンス（35.3%）の順に高かった。経済学・経営学、免疫学、社会科学、数学は、「はい」を選ぶ回答者がいなかった。

[4] データ公開の障壁の程度

分野によるカレントデータの管理・公開資源の充足度に差があるかどうかを調べた結果、データ公開用リポジトリと研究中のストレージに差がみられた。図 3-8 に、分野別データ公開用リポジトリの充足度を示す。排列は、「不十分」と「やや不十分」の合計が多い順とした（以下、同様とする）。

精神医学・心理学は「不十分」と「やや不十分」の合計が92.3%と全分野の中で最も高く、「充分」、「ほぼ充分」を選択した回答者はいなかった。臨床医学は、「不十分」と「やや不十分」の合計は59.1%で全体の15位であったが、「充分」を選択した回答者はいなかった。植物・動物学分野は「充分」は26.3%、「充分」と「ほぼ充分」の合計は42.1%で、いずれも全分野の中で最も高かった。コンピュータサイエンスは「充分」は12.7%、「充分」と「ほぼ充分」の合計は34.9%で、いずれも全分野の中で2位であった。

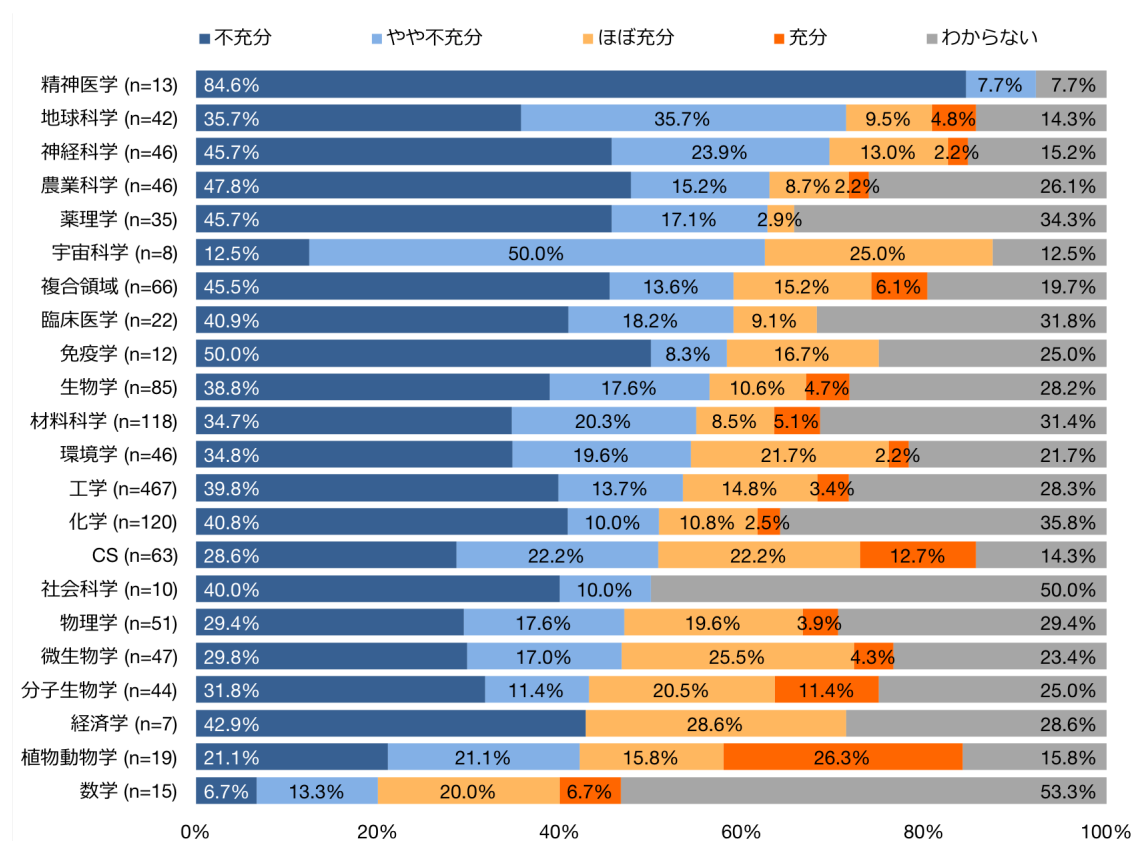


図 3-8 分野別「データ公開用リポジトリ」の充足度 (n=1,382)

図 3-9 に、分野別「研究中のストレージ」の充足度を示す。宇宙科学は「不十分」と「やや不十分」の合計が 87.5%と全分野の中で最も高く、「わからない」を選択した回答者はいなかった。精神医学・心理学は「不十分」と「やや不十分」の合計が 76.9%であり、「充分」を選択した回答者は全くいなかった。充足度が高かった分野はコンピュータサイエンスと植物・動物学であった。コンピュータサイエンスは「充分」は 25.4%, 「充分」と「ほぼ充分」の合計は 61.9%であり、いずれも全分野の中で最も高かった。植物・動物学分野は「充分」は 21.1%, 「充分」と「ほぼ充分」の合計は 57.9%であり、いずれも全分野の中で 2 位であった。

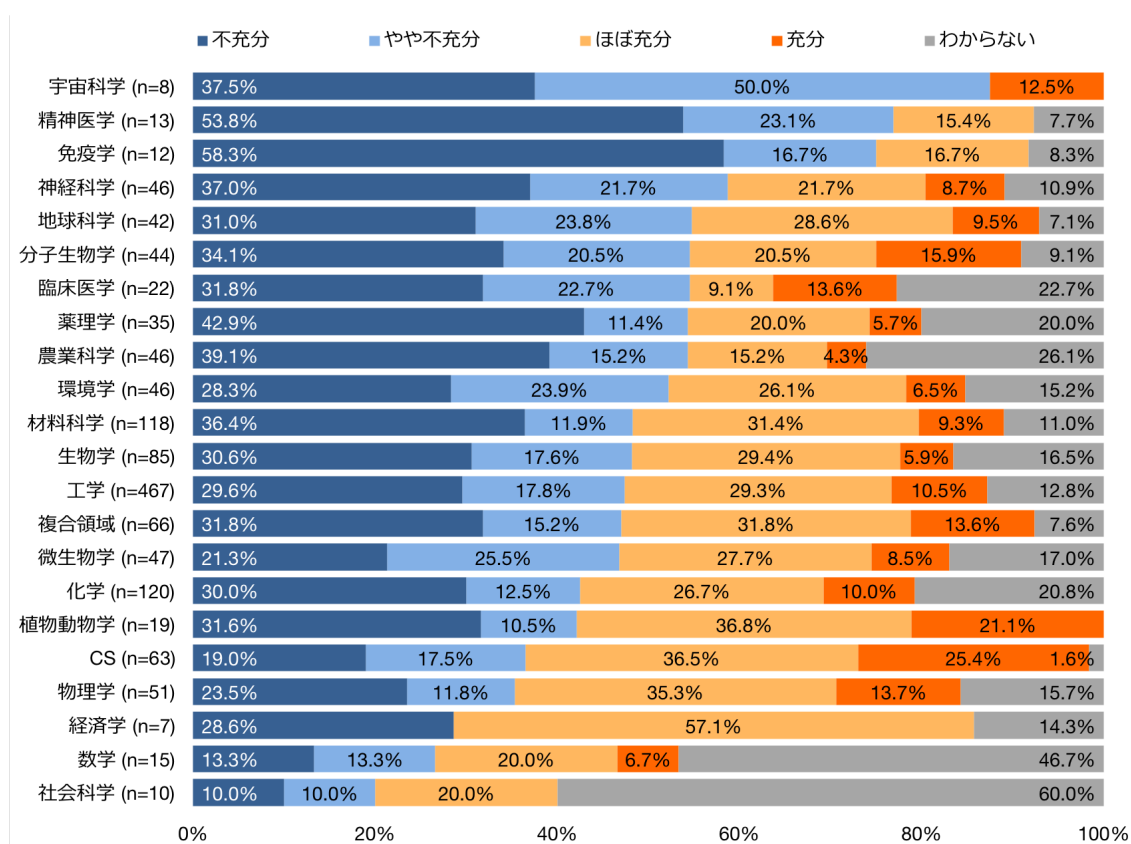


図 3-9 分野別「研究中のストレージ」の充足度 (n=1,382)

3 章 日本の研究者によるデータ公開状況と内的要因

分野によってカレントデータを公開する場合の懸念に差があるかどうか調べた結果、商用利用される可能性、機密・プライバシー情報、引用せずに利用される可能性、先に論文を出版される可能性、研究の誤りを発見される可能性に差がみられた。

図 3-10 に、商用利用される可能性への懸念の強さを示す。排列は、「問題である」と「やや問題である」の合計が多い順とした（以下、同様とする）。

「問題である」と「やや問題である」の選択率が高かったのは、免疫学（91.6%）、臨床医学（81.3%）であり、「問題ではない」を選択した回答者はいなかった。また、免疫学、経済学・経営学、宇宙科学の回答者で「わからない」を選択した回答者はいなかった。選択率が低かったのは、植物・動物学分野（31.6%）、宇宙科学（37.5%）、社会科学（40.0%）であった。

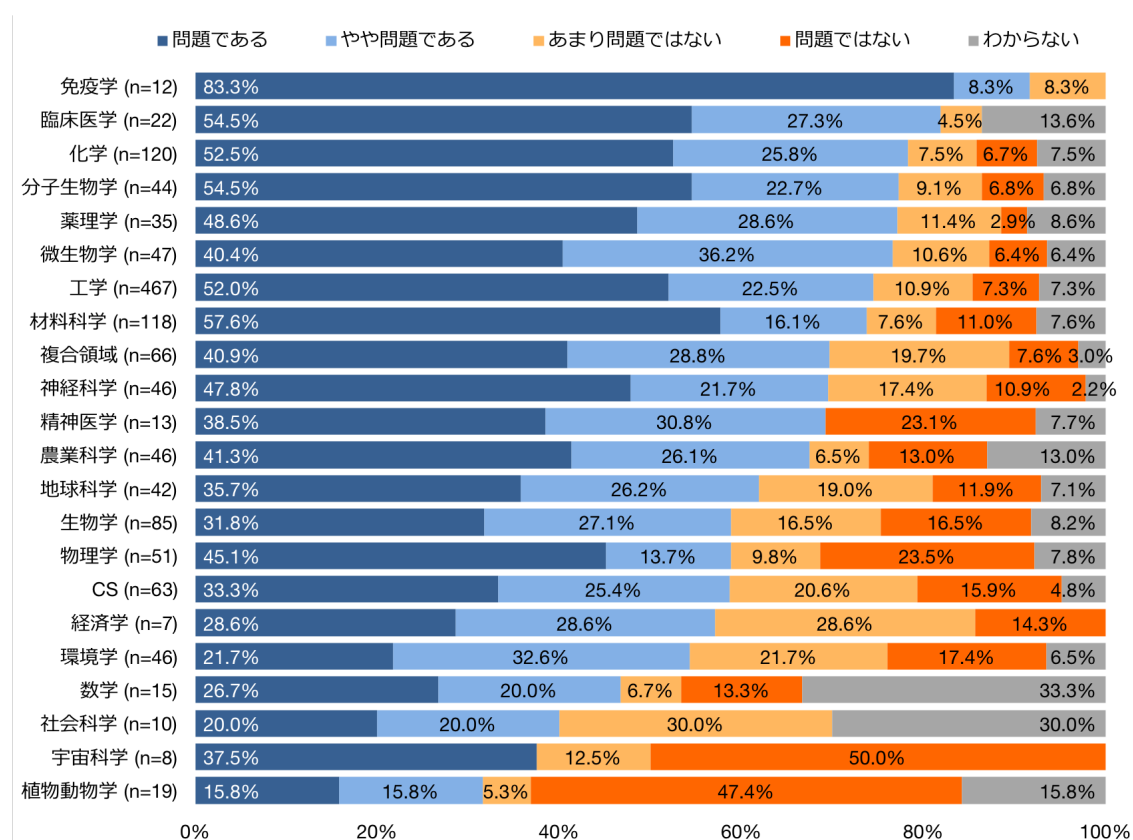


図 3-10 分野別「商用利用される可能性」への懸念 (n=1,382)

図 3-11 に、機密・プライバシー情報への懸念の強さを示す。臨床医学は、「問題である」と「やや問題である」の合計は 90.9%で全体の 1 位であり、かつ、「問題ではない」や「あまり問題ではない」を選択した回答者はいなかった。次いで複合領域 (86.3%), 免疫学 (83.3%) の順に多かった。精神医学・心理学は「問題である」を選択した回答者が 76.9%であり、全分野の 1 位であった。「やや問題である」を選んだ回答者はおらず、「問題ではない」や「あまり問題ではない」の合計は 23.1%, 「わからない」を選択した回答者はいなかった。このほか、免疫学, 社会科学, 経済学・経営学, 宇宙科学の回答者で「わからない」を選択した回答者はいなかった。選択率が低かったのは、数学 (26.7%), 宇宙科学 (37.5%), 植物・動物学 (42.1%) であった。

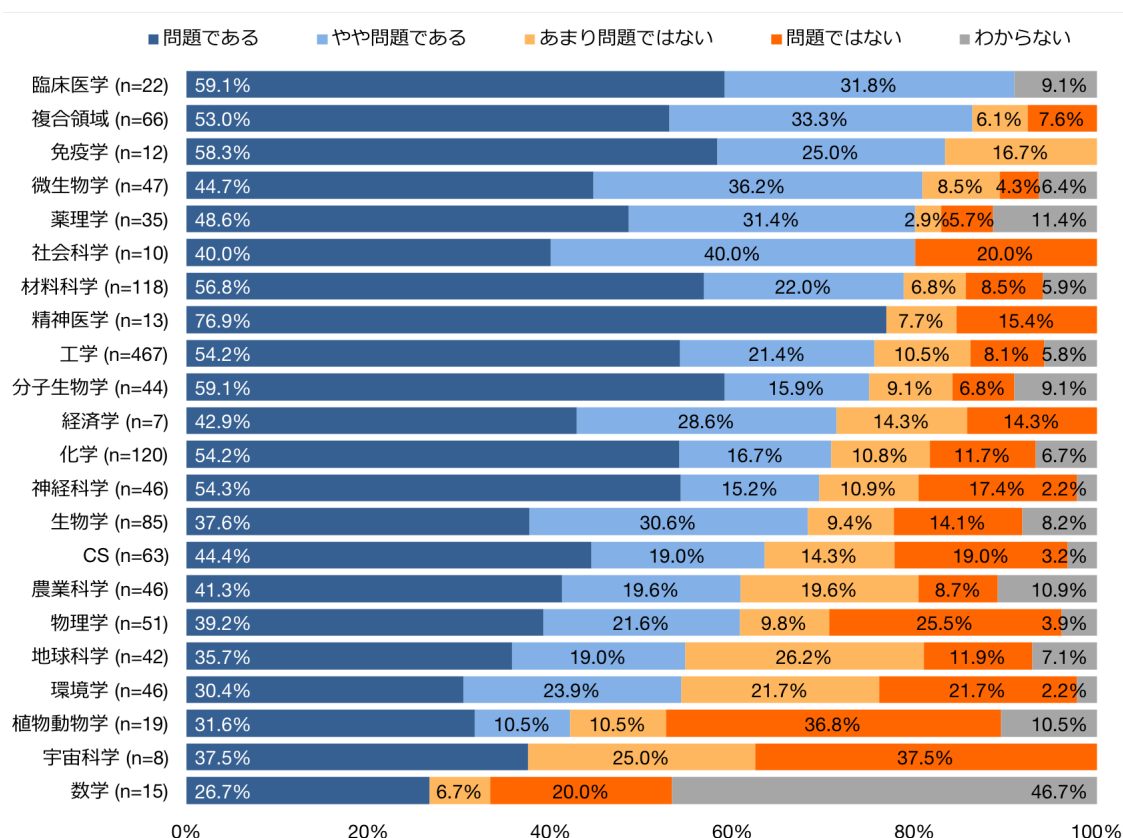


図 3-11 分野別「機密・プライバシー情報」への懸念 (n=1,382)

3 章 日本の研究者によるデータ公開状況と内的要因

図 3-12 に、引用せずに利用される可能性への懸念の強さを示す。全体的に懸念が強く、経済学は全ての回答者が「問題である」または「やや問題である」を選択していた。次いで材料科学 (92.4%)、精神医学・心理学 (92.3%) であり、選択率が低かったのは数学 (53.4%)、宇宙科学 (62.5%)、臨床医学 (77.2%) であった。

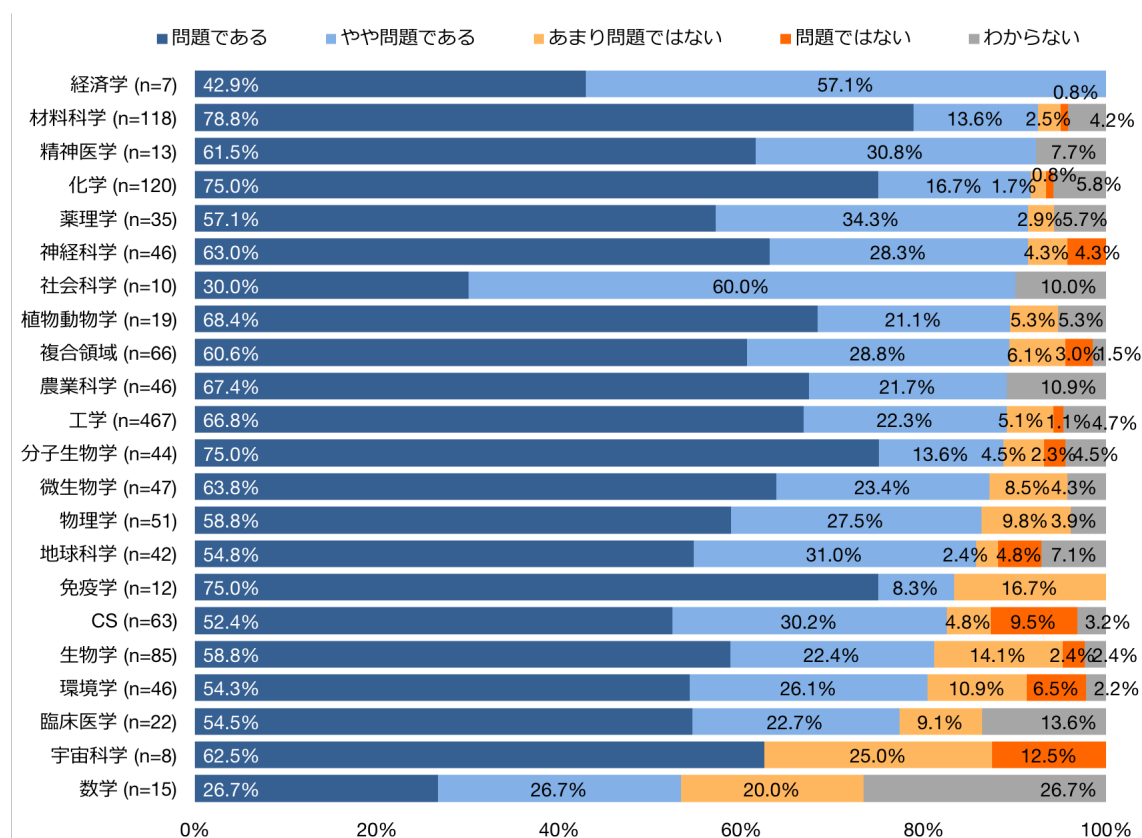


図 3-12 分野別「引用せずに利用される可能性」への懸念 (n=1,382)

図 3-13 に、先に論文を出版される可能性への懸念の強さを示す。引用せずに利用される可能性と同様、全体的に懸念が強かった。免疫学は全ての回答者が「問題である」を選択しており、2位の植物・動物学は「問題である」と「やや問題である」の合計が94.8%であった。選択率が低かったのは、数学（46.6%）、経済学・経営学（71.4%）、コンピュータサイエンス（74.6%）であった。

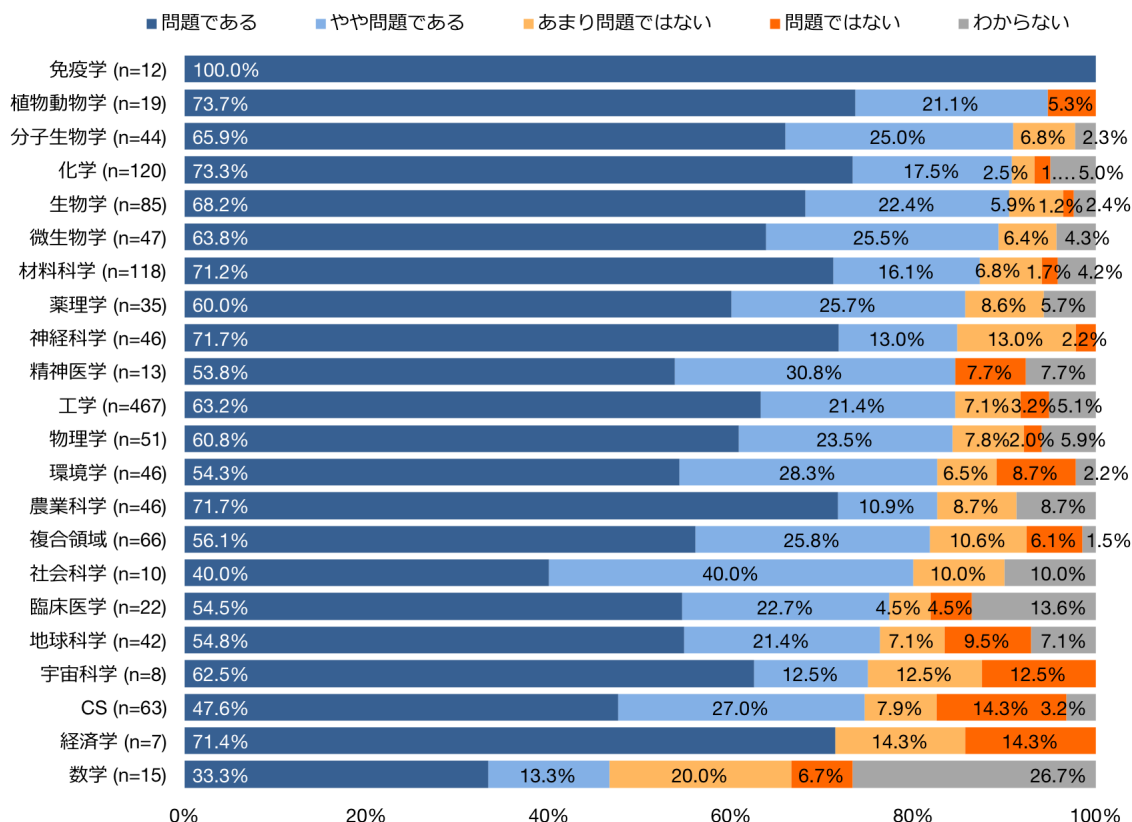


図 3-13 分野別「先に論文を出版される可能性」への懸念 (n=1,382)

3 章 日本の研究者によるデータ公開状況と内的要因

図 3-14 に、研究の誤りを発見される可能性への懸念の強さを示す。この質問は全体的に懸念が弱く、社会科学と宇宙科学は「問題である」や「やや問題である」を選択した回答者はいなかった。比較的懸念が強かったのは、免疫学（「問題である」と「やや問題である」の合計 26.5%）、数学（20.0%）、微生物学（19.1%）であった。

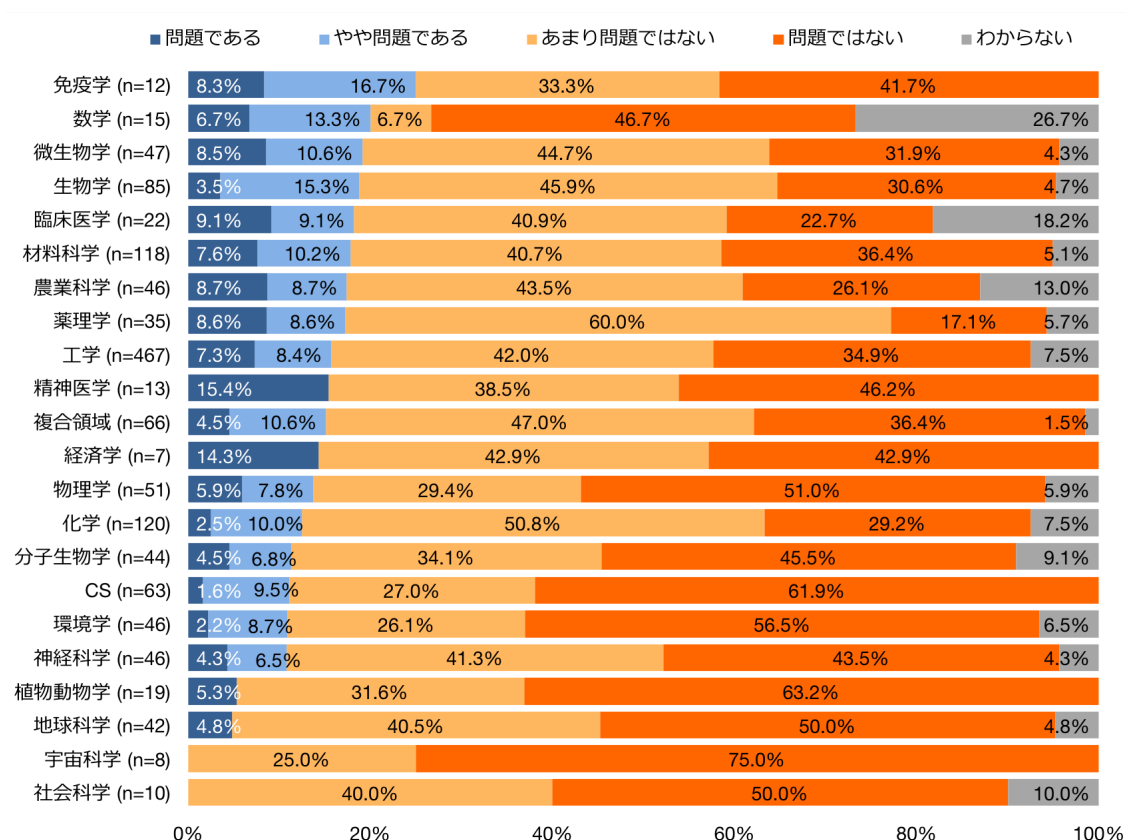


図 3-14 分野別「研究の誤りを発見される可能性」への懸念 (n=1,382)

分野別の「機密・プライバシー情報」と「商用利用される可能性」について、「問題である」と「やや問題である」の選択率の合計の相関を調べたところ、Pearson の相関係数は $r=0.732$ であった ($p<0.01$)。図 3-15 に散布図を示す。図では、数値が大きいほど懸念が強いことを示している。商用利用される可能性への懸念が強い分野は、機密・プライバシー情報への懸念も強い傾向が確認された。

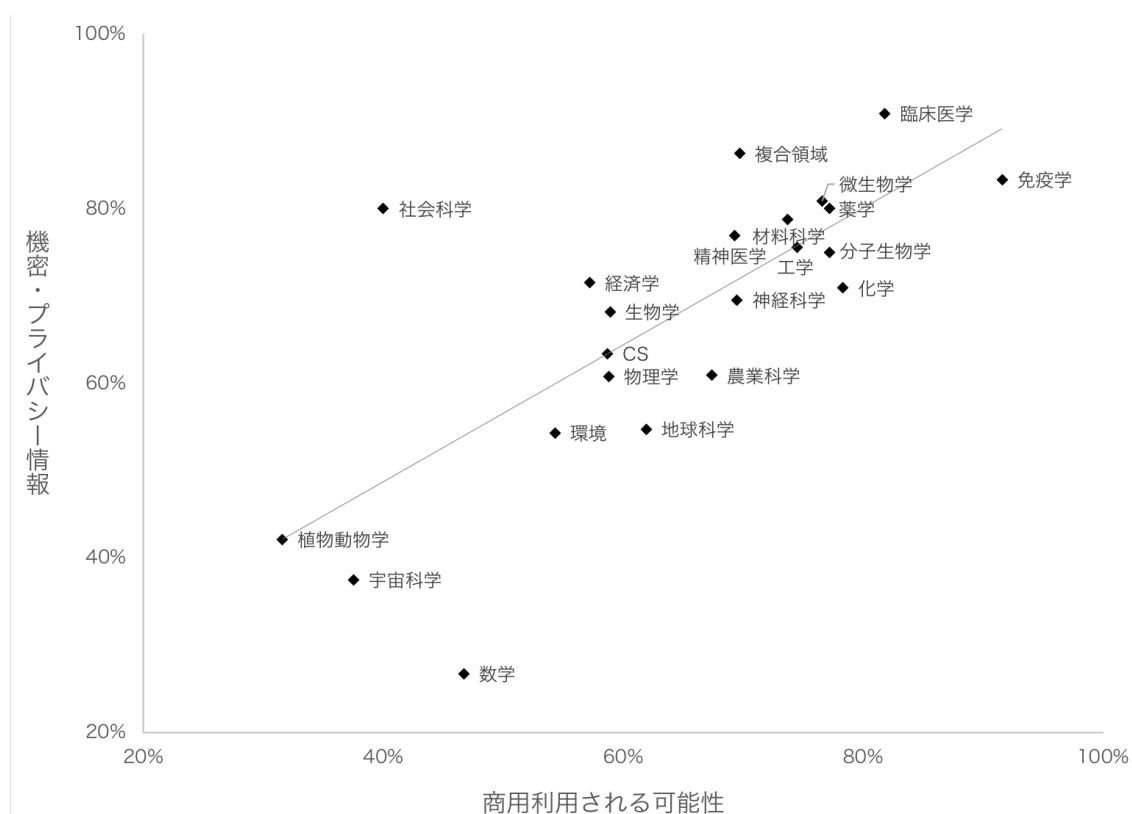


図 3-15 分野別データ公開の懸念（機密と商用利用）

3 章 日本の研究者によるデータ公開状況と内的要因

分野による「データを公開した場合の理解の難しさ」に対する認識は、同じ分野の研究者についてはあまり差がみられなかった（付表 47）。異分野の研究者について、図 3-16 に分野ごとの結果を示す。排列は「難しいと思う」と「やや難しいと思う」の合計が多い順とした。図 3-16 によれば、「難しいと思う」は 14.3%から 66.7%まで、「やや難しいと思う」は 25.0%から 53.8%まで差がみられた。なお、データ公開経験との関連はみられなかった。つまり、難しいと認識している分野の回答者ほどデータ公開率が低いといったことはなかった。

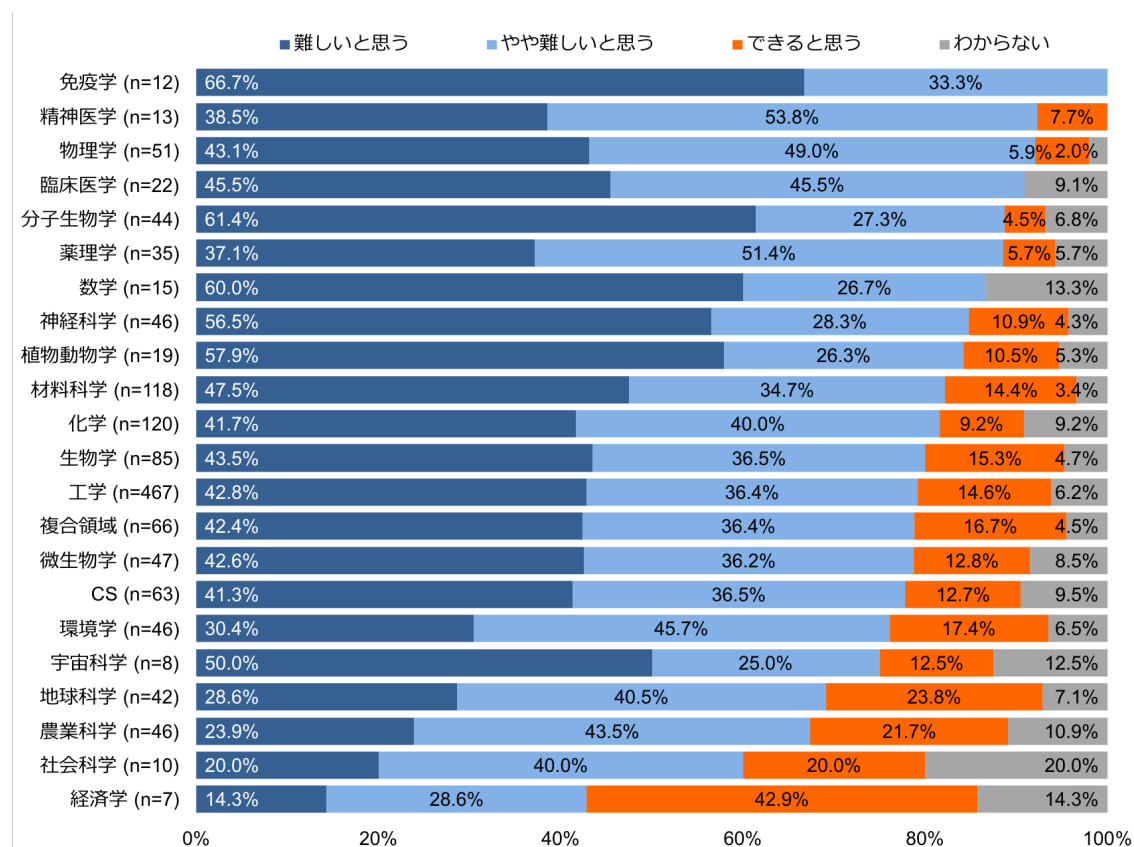


図 3-16 分野別異分野の研究者によるデータの理解の難しさ (n=1,382)

「難しいと思う」と「やや難しいと思う」の合計が高い順に、免疫学 (100.0%)、精神医学・心理学 (92.3%)、物理学 (92.1%) であった。免疫学と精神医学・心理学で「わからない」を選択した回答者はいなかった。選択率の合計が低かったのは経済学・経営学 (42.9%)、社会科学 (60.0%)、農業科学 (67.4%) であった。

[5] データの提供と被提供状況

表 3-28 に、分野別のデータの提供と被提供状況について、「よくある」と「たまにある」の選択率の合計を示す。排列は提供経験の値が高い順として、順位を付した。また、参考の

ためデータ公開率と順位を示した。

表 3-28 分野別データ提供・被提供経験

分野	順位	提供経験※	順位	被提供経験※	順位	データ公開率
宇宙科学	1	87.5%	3	62.5%	15	50.0%
数学	2	66.7%	4	60.0%	17	46.7%
地球科学	3	66.6%	1	66.6%	12	52.4%
環境	4	60.0%	6	51.1%	3	69.6%
植物動物学	5	57.9%	2	63.2%	1	84.2%
CS	6	52.4%	5	55.5%	6	65.1%
分子生物学	7	52.3%	7	50.0%	2	70.5%
農業科学	8	50.0%	17	32.6%	8	56.5%
複合領域	8	50.0%	9	44.0%	9	54.5%
材料科学	10	46.6%	12	39.0%	18	45.8%
工学	11	43.7%	13	37.9%	21	41.1%
物理学	12	43.2%	16	33.3%	10	52.9%
経済学	13	42.9%	10	42.9%	20	42.9%
免疫学	14	41.7%	7	50.0%	4	66.7%
社会科学	15	40.0%	11	40.0%	15	50.0%
神経科学	16	39.1%	19	28.3%	13	52.2%
生物学	17	38.8%	14	34.2%	7	61.2%
臨床医学	18	36.4%	20	27.3%	19	45.5%
微生物学	19	35.5%	15	33.4%	4	66.7%
化学	20	30.6%	18	29.7%	10	52.9%
精神医学	21	23.1%	21	15.4%	22	30.8%
薬理学	22	20.0%	22	14.3%	14	51.4%
平均		44.2%		38.9%		51.2%
回答者数		1,383		1,383		1,384

※「よくある」と「たまにある」の選択率の合計

分野別の提供経験と被提供経験は分野によるばらつきがみられた。また、分野別の提供経験と被提供経験の関連を調べたところ、Pearson の相関係数は $r=0.881$ であった ($p<0.001$)。提供よりも被提供の順位が高い分野は、地球科学、植物・動物学、コンピュータサイエンスなどであった。データ公開率と提供経験は関連がみられず、被提供経験とは中程度の相関がみられた (Pearson の相関係数 $r=0.455$, $p<0.05$)。被提供の上位 3 分野を確認すると、地球科

3 章 日本の研究者によるデータ公開状況と内的要因

学のデータ公開率は 12 位 (52.4%), 植物・動物学は 1 位 (84.2%), 宇宙科学は 16 位 (50.0%), であった。

[6] 公開データの入手経験

表 3-24 に示した公開データの入手経験は, 分野によるばらつきがみられた。入手経験がある回答者の比率が高い順に, 分子生物学・遺伝学 (97.7%), 地球科学 (90.5%), 植物・動物学 (89.5%) であり, 低い順に, 精神医学・心理学 (46.2%), 臨床医学 (50.0%), 工学 (70.0%) であった。

図 3-17 に分野別の公開データの入手経験がある回答者の比率とデータ公開経験がある回答者の比率の散布図を示す。両者には関連がみられ, 公開データの入手経験がある回答者の比率が高い分野は, データ公開経験がある回答者の比率も高かった (Pearson の相関係数 $r=0.658$, $p<0.01$)。

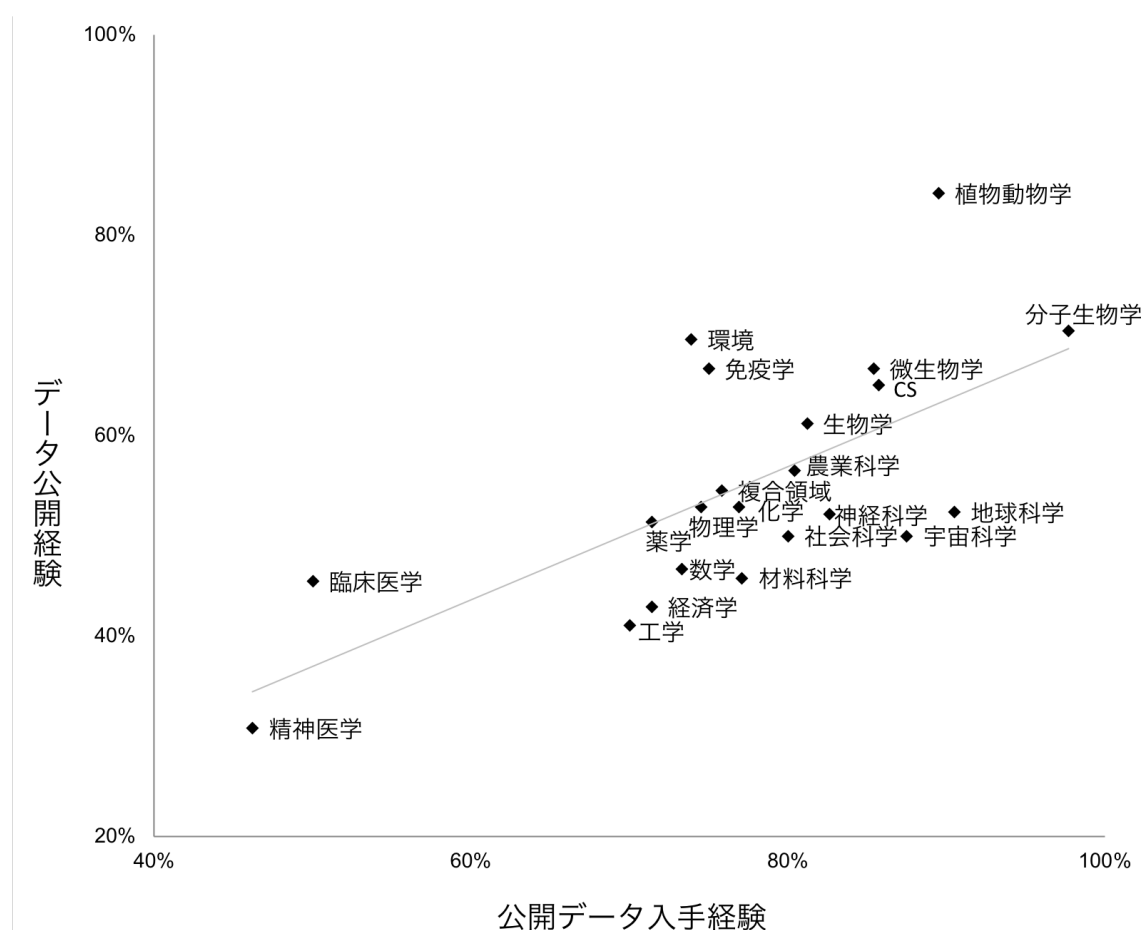


図 3-17 分野別公開データの入手経験とデータ公開経験の関連

公開データの利用目的として尋ねた「自身の研究のアイデアや仮説の参考にする (以下,

「研究の参考」), 「再分析・再利用して自身の研究を行う (以下, 「研究に再利用」), 「(他者の) 研究を再現・追試する (以下, 「研究を再現」)」頻度について, 「研究に再利用」は分野による差がみられた。図 3-18 に, 「行わない」頻度が低い順に集計結果を示す。「行う」または「たまに行う」比率が高い順に, 経済学・経営学 (合計 80.0%), 地球科学 (78.9%), コンピュータサイエンス (75.9%) であった。比率が低かったのは, 精神医学・心理学 (0%), 臨床医学 (9.1%), 社会科学 (37.5%) であった。

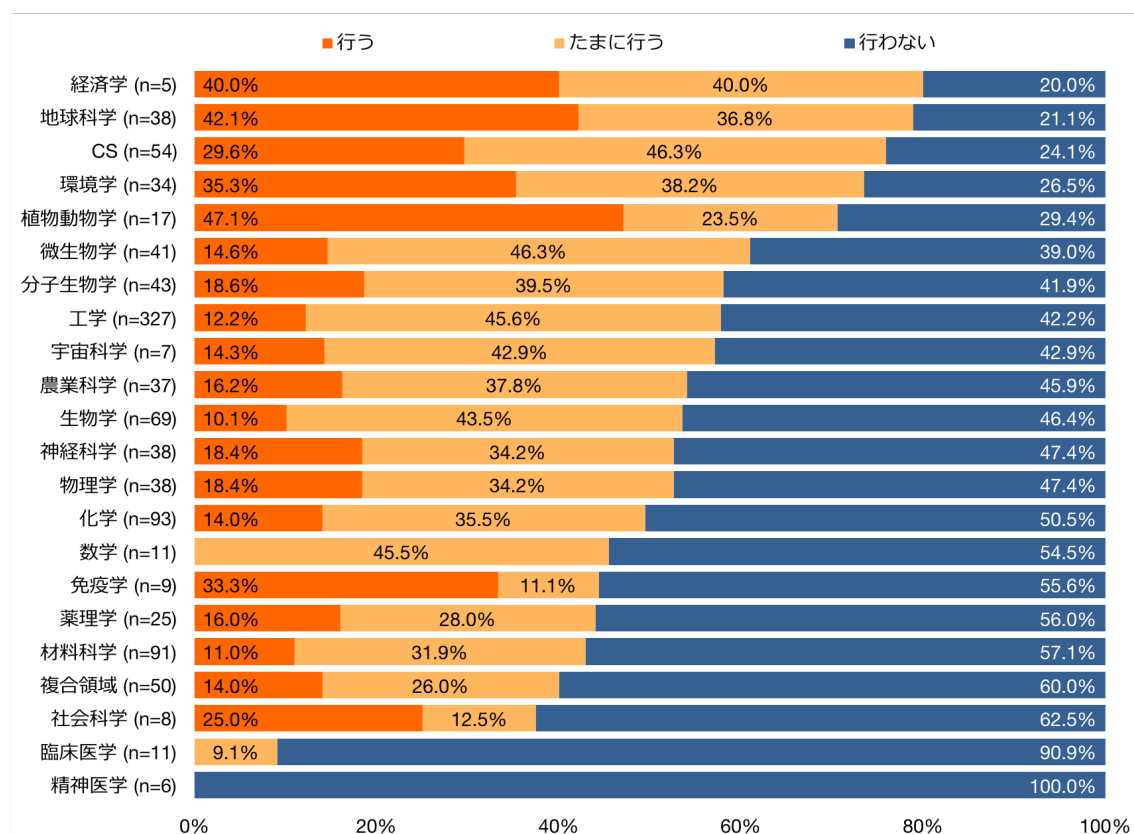


図 3-18 分野別公開データの再分析・再利用を行う頻度 (n=1,052)

公開データの入手経験がある回答者の比率とデータの公開経験がある回答者の比率 (以下, 「公開率」) に関連がみられたことから, 分野別にデータリポジトリの利用 (検索, 入手, 公開) 状況を確認した。表 3-29 に「データ公開率」, 「論文／検索」(論文の検索ツールとして「論文やプレプリントのサーバー (PubMed Central, arXiv, J-Stage など)」を選択した回答者の比率), 「データリポジトリ」の「検索」(データの検索ツールとして「特定分野のデータリポジトリ (DDBJ や ICPSR など)」を選択した回答者の比率), 「入手」(公開データの入手先としてデータリポジトリを選択した回答者の比率), 「公開」(データの公開先としてデータリポジトリを選択した回答者の比率) をそれぞれ示す。排列は, データ公開率が高い分野順とした。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-29 分野別リポジトリの利用状況

分野	データ 公開率	論文	データリポジトリ		
		検索	検索	入手	公開
植物動物学	84.2%	68.4%	11.8%	68.4%	52.6%
分子生物学	70.5%	81.8%	7.1%	56.8%	29.5%
環境学	69.6%	22.2%	0.0%	19.6%	19.6%
免疫学	66.7%	66.7%	0.0%	33.3%	0.0%
微生物学	66.7%	75.0%	7.3%	37.5%	27.1%
CS	65.1%	49.2%	3.8%	25.4%	9.5%
生物学	61.2%	89.4%	4.5%	34.1%	18.8%
農業科学	56.5%	55.6%	0.0%	26.1%	13.0%
複合領域	54.5%	54.5%	2.1%	21.2%	10.6%
化学	52.9%	40.5%	0.0%	6.6%	1.7%
物理学	52.9%	42.0%	0.0%	13.7%	7.8%
地球科学	52.4%	36.6%	0.0%	16.7%	4.8%
神経科学	52.2%	76.1%	2.8%	15.2%	13.0%
薬理学	51.4%	74.3%	0.0%	5.7%	5.7%
社会科学	50.0%	20.0%	12.5%	10.0%	10.0%
宇宙科学	50.0%	62.5%	0.0%	0.0%	12.5%
数学	46.7%	80.0%	0.0%	13.3%	0.0%
材料科学	45.8%	37.3%	2.2%	9.3%	1.7%
臨床医学	45.5%	86.4%	0.0%	4.5%	0.0%
経済学	42.9%	57.1%	20.0%	0.0%	0.0%
工学	41.1%	33.0%	2.5%	11.8%	3.2%
精神医学	30.8%	69.2%	0.0%	15.4%	7.7%
全体平均	51.2%	48.3%	2.6%	17.6%	8.4%
回答者数	1,384	1,380	1,025	1,384	1,384

検索ツールとして、論文のサーバーは生物学・生化学の 89.4%から社会科学の 20.0%まで、22 分野の回答者がそれぞれ複数名ずつ選んでいた。一方、特定分野のデータリポジトリ（以下、「分野リポジトリ」）を選んでいたのは、工学（8 名）、生物学・生化学、微生物学、分子生物学・遺伝学（各 3 名）など 11 分野の回答者のみであり、その他の 11 分野は選択する回答者がいなかった。

表 3-29 の「入手」をみると、データの入手先として分野リポジトリを選択した回答者の比

率は、微生物学 (37.5%)、生物学・生化学 (34.1%)、免疫学 (33.3%) の順に多く、宇宙科学と経済学は 0%であった。データの入手先として分野リポジトリを選択した回答者の比率とデータ公開率の相関を調べたところ、Pearson の相関係数は $r=0.814$ であった ($p<0.001$)。また、公開先として分野リポジトリを選択した回答者の比率とデータの公開率の相関を調べたところ、Pearson の相関係数は $r=0.761$ であった ($p<0.001$)。つまり、データ公開率が高い分野では、データの入手先や公開先として分野リポジトリが利用される傾向があることが確認された。

[7] データのプロファイル

カレントデータのプロファイルについて、分野によって差があるかどうかを確認した。まず、表 3-30 に分野別のカレントデータの量を示す。排列はデータ量が多い順とした。宇宙科学、地球科学、分子生物学・遺伝学の順に多く、経済学・経営学、精神医学・心理学、農業科学の順にデータ量が少なかった。分野によってデータ量には差がみられたものの、データ量とデータ公開率には関連がみられなかった。つまり、データの量が多い分野はデータ公開率が低いといった関連はみられなかった。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-30 分野別カレントデータの量

分野	MB 以下		GB		TB		PB 以上		合計	
	人	%	人	%	人	%	人	%	人	%
宇宙科学	0	0.0	4	50.0	4	50.0	0	0.0	8	100.0
地球科学	7	17.1	22	53.7	12	29.3	0	0.0	41	100.0
分子生物学	5	12.5	28	70.0	7	17.5	0	0.0	40	100.0
複合領域	10	17.2	37	63.8	11	19.0	0	0.0	58	100.0
微生物学	5	13.2	28	73.7	5	13.2	0	0.0	38	100.0
物理学	8	17.0	32	68.1	6	12.8	1	2.1	47	100.0
生物学	11	14.3	56	72.7	10	13.0	0	0.0	77	100.0
工学	80	19.5	257	62.7	72	17.6	1	0.2	410	100.0
CS	15	26.3	28	49.1	14	24.6	0	0.0	57	100.0
材料科学	16	14.7	81	74.3	12	11.0	0	0.0	109	100.0
神経科学	11	25.6	23	53.5	9	20.9	0	0.0	43	100.0
化学	21	20.4	69	67.0	12	11.7	1	1.0	103	100.0
免疫学	2	20.0	7	70.0	1	10.0	0	0.0	10	100.0
臨床医学	4	23.5	11	64.7	2	11.8	0	0.0	17	100.0
薬理学	7	21.9	22	68.8	3	9.4	0	0.0	32	100.0
数学	4	40.0	4	40.0	2	20.0	0	0.0	10	100.0
社会科学	1	20.0	4	80.0	0	0.0	0	0.0	5	100.0
植物動物学	4	21.1	15	78.9	0	0.0	0	0.0	19	100.0
環境学	16	39.0	19	46.3	6	14.6	0	0.0	41	100.0
農業科学	16	39.0	24	58.5	0	0.0	1	2.4	41	100.0
精神医学	6	50.0	6	50.0	0	0.0	0	0.0	12	100.0
経済学	5	83.3	1	16.7	0	0.0	0	0.0	6	100.0
合計	254	20.8	778	63.6	188	15.4	4	0.3	1,224	100.0

カレントデータに機密情報が含まれる回答者の比率は、分野によって差がみられた。表 3-24 に示した機密情報「あり」の比率が高い分野は免疫学(75.0%), 精神医学・心理学(69.2%), 臨床医学(68.2%)であり、「なし」の比率が高い分野は植物・動物学(84.2%), 数学(66.7%), 宇宙科学(62.5%)であった。これらは図 3-11 に示した「機密・プライバシー情報」への懸念が強い・弱い3分野とおおむね一致していた。また、分野別に「機密・プライバシー情報」について「問題である」と「やや問題である」の選択率の合計と機密情報「あり」の比率の相関を調べたところ、Pearson の相関係数は $r=0.775$ であった($p<0.001$)。同様に、「商用利用

される可能性」への懸念の強さとの相関を調べた結果、Pearson の相関係数は $r=0.837$ であった ($p<0.001$)。つまり、カレントデータに機密情報が含まれる分野の回答者は、データを公開する場合の機密・プライバシー情報や商用利用される可能性への懸念が強いことがわかった。

3.2.9 所属機関や年齢層による差の比較

3.2.1 から 3.2.7 までの項目について、属性（所属機関と年齢層）ごとに集計した。以下では所属や年齢層によって差がみられた項目をそれぞれ示す。また、次節（3.3 考察）の表 3-61 に、所属や年齢層によって差がみられた項目の一覧を示す。

[1] 所属機関による差

表 3-31 に所属機関別のデータ公開経験を示す。大学（56.9%）、公的機関・団体（55.6%）、企業（32.4%）の順に公開経験がある回答者の比率が高かった。

表 3-31 所属機関別データ公開経験

所属	あり	なし	わからない	使わない	回答者数
大学	480	353	10	0	843
	56.9%	41.9%	1.2%	0.0%	100.0%
公的機関・団体	124	94	5	0	223
	55.6%	42.2%	2.2%	0.0%	100.0%
企業	105	205	12	2	324
	32.4%	63.3%	3.7%	0.6%	100.0%
合計	709	652	27	2	1,390
	51.0%	46.9%	1.9%	0.1%	100.0%

3 章 日本の研究者によるデータ公開状況と内的要因

図 3-19 に、データ整備・公開のための人材の所属機関別の結果を示す。大学の回答者は不足感が最も強く、「不十分」、「やや不十分」の合計は、82.7%であった。公的機関・団体は 77.2%，企業は 70.8%であった。企業は「わからない」の比率も高かった（24.2%）。

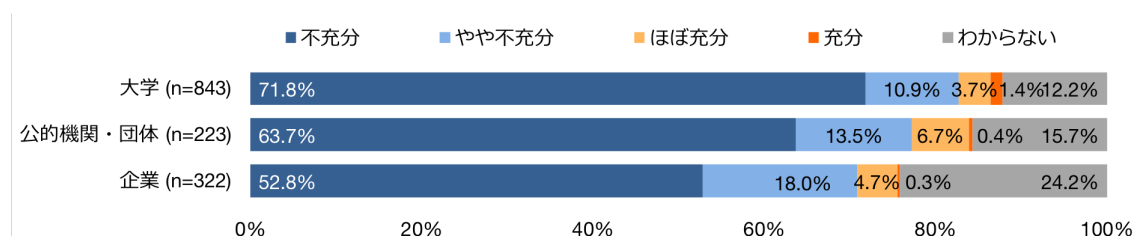


図 3-19 所属機関別「データ整備・公開のための人材」の充足度 (n=1,388)

図 3-20 に、データ整備・公開のための時間の所属機関別の結果を示す。大学の回答者は不足感が最も強く、「不十分」、「やや不十分」の合計は 77.0%であった。公的機関・団体は 73.5%，企業は 68.0%であった。企業は「わからない」の比率も高かった（23.6%）。

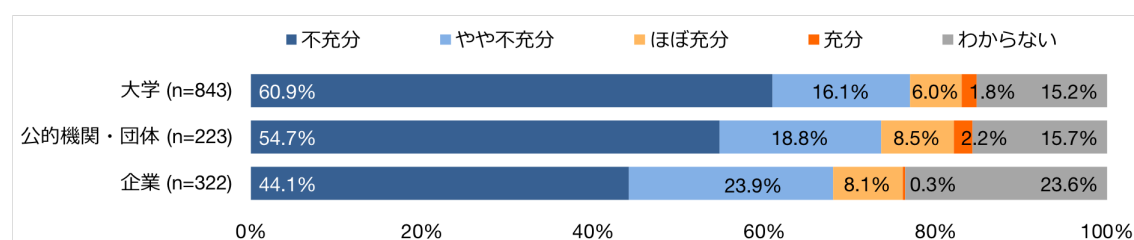


図 3-20 所属機関別「データ整備・公開のための時間」の充足度 (n=1,388)

図 3-21 に、データ整備・公開のための資金の所属機関別の結果を示す。大学の回答者は不足感が最も強く、「不十分」、「やや不十分」の合計は 80.0%であった。公的機関・団体は 73.5%，企業は 58.7%であった。

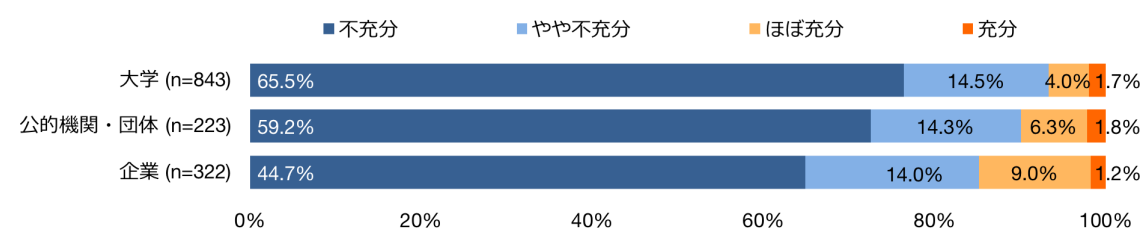


図 3-21 所属機関別「データ整備・公開のための資金」の充足度 (n=1,388)

図 3-22 に、商用利用される可能性への懸念の所属機関別の結果を示す。企業の回答者の懸念が最も強く、「問題である」、「やや問題である」の合計は 78.2%であった。大学や公的機関・団体の回答者の懸念も強く、大学は 69.1%，公的機関・団体は 61.4%であった。

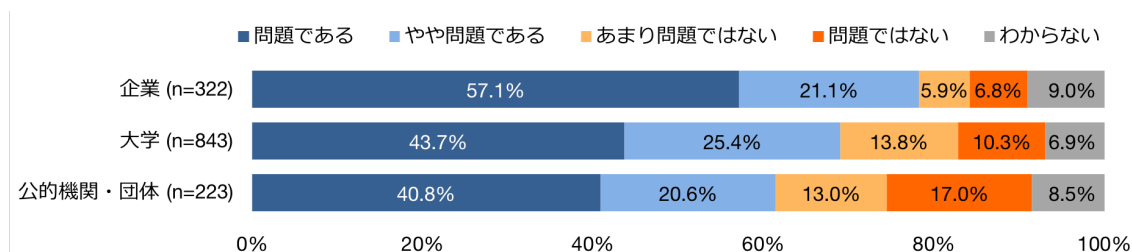


図 3-22 所属機関別「商用利用される可能性」への懸念 (n=1,388)

図 3-23 に機密・プライバシー情報への懸念の所属機関別の結果を示す。商用利用と同じく、企業の回答者の懸念が最も強く、「問題である」、「やや問題である」の合計は 78.5%であった。自由回答では、特許のため公開は困難であるという意見が多数みられた。大学や公的機関・団体の回答者の懸念は商用利用よりもやや強く、大学は 70.7%，公的機関・団体は 66.8%であった。

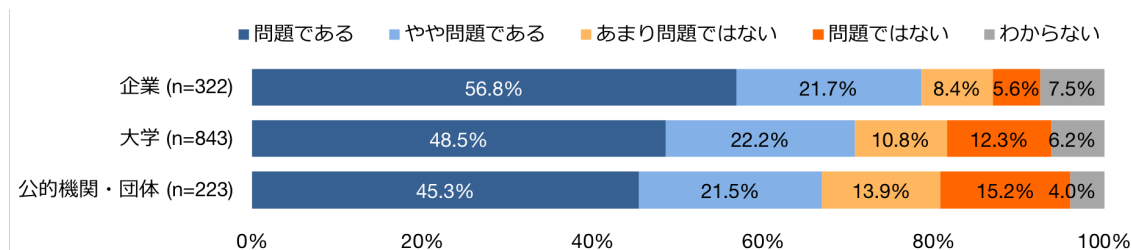


図 3-23 所属機関別「機密・プライバシー情報」への懸念 (n=1,388)

3 章 日本の研究者によるデータ公開状況と内的要因

図 3-24 に、引用せずに利用される可能性への懸念の強さを所属機関別に示す。大学の回答者の懸念が最も強く、「問題である」、「やや問題である」の合計は 89.0%であった。次いで公的機関・団体（87.9%）、企業（84.2%）の順であった。

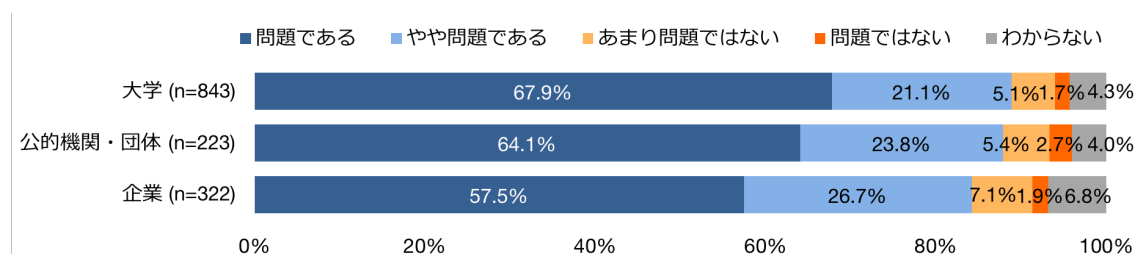


図 3-24 所属機関別「引用せずに利用される可能性」への懸念 (n=1,388)

図 3-25 に、先に論文を出版される可能性への懸念の強さを所属機関別に示す。引用せずに利用される可能性と同様に、大学の回答者の懸念が最も強く、「問題である」、「やや問題である」の合計は 87.5%であった。「問題である」とする回答者は公的機関・団体、企業の順であったが、「問題である」、「やや問題である」の合計は、公的機関・団体が 79.8%、企業が 80.4%であり、わずかに逆転が起きていた。「問題ではない」を選択した回答者の比率は、公的機関・団体が 6.3%で最も高かった。

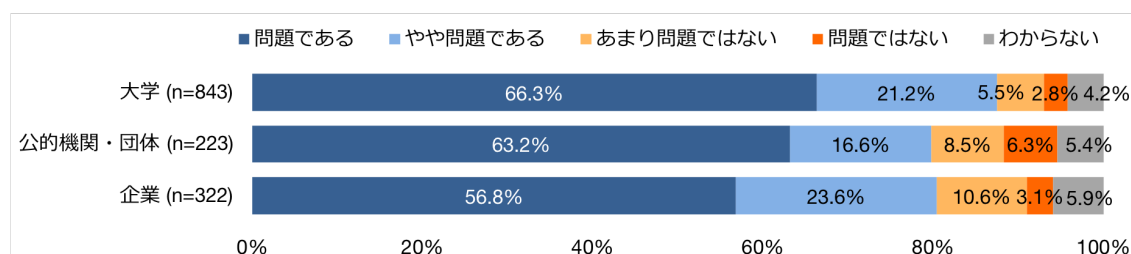


図 3-25 所属機関別「先に論文を出版される可能性」への懸念 (n=1,388)

図 3-26 に、研究の誤りを発見される可能性への懸念の強さを所属機関別に示す。全体的に懸念は弱かったが、企業の回答者の懸念が最も強く、「問題である」、「やや問題である」の合計は 16.4%であった。次いで大学（15.3%）、公的機関・団体（9.8%）の順であった。

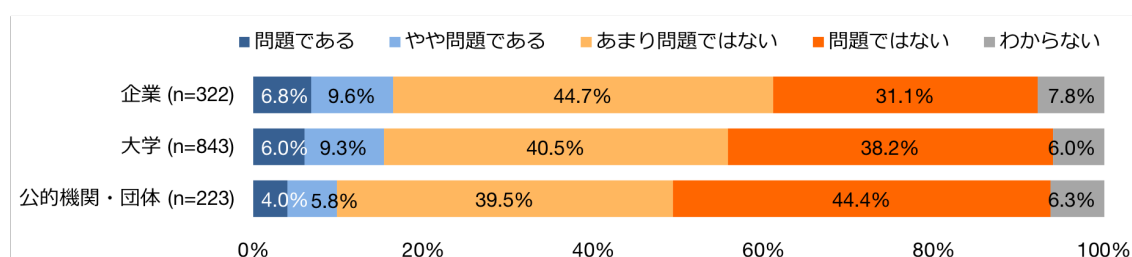


図 3-26 所属機関別「研究の誤りを発見される可能性」への懸念 (n=1,388)

データを公開する場合の懸念について尋ねた項目のうち、誤解や誤用の可能性は所属機関による差はあまりなかった。クロス集計表を付表 48 に示す。

表 3-32 に所属機関別カレントデータの所有権をもつ人や組織の有無を示す。「あり」の比率は企業が 88.8%、公的機関・団体が 86.5%、大学が 80.9%であった。クロス集計の結果、企業の回答者の 26 名 (8.1%) が「所属機関以外の研究機関」を、大学の 41 名 (4.9%)、公的機関・団体の 12 名 (5.4%) が「所属機関以外の企業」を選択しており、所有権が産学にわたるデータが存在することが確認された。

表 3-32 所属機関別カレントデータの所有権をもつ人・組織

所属	所有権		あり		なし		わからない		回答者数	
	人数	比率	人数	比率	人数	比率	人数	比率	人数	比率
企業	286	88.8%	14	4.3%	22	6.8%	322	100.0%		
公的機関・団体	193	86.5%	16	7.2%	14	6.3%	223	100.0%		
大学	681	80.9%	84	10.0%	77	9.1%	842	100.0%		
合計	1,160	83.6%	114	8.2%	113	8.1%	1,387	100.0%		

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-33 に所属機関別の機密情報の有無を示す。所属別の「あり」の比率は、企業が 70.5%、公的機関・団体が 45.2%、大学が 42.0%であった。

表 3-33 所属機関別機密情報の有無

所属	機密		あり		なし		わからない		回答者数	
	人数	比率	人数	比率	人数	比率	人数	比率	人数	比率
企業	225	70.5%	54	16.9%	40	12.5%	319	100.0%		
公的機関・団体	100	45.2%	88	39.8%	33	14.9%	221	100.0%		
大学	351	42.0%	351	42.0%	133	15.9%	835	100.0%		
合計	676	49.2%	493	35.9%	206	15.0%	1,375	100.0%		

[2] 年齢による差

年齢による差を明らかにするために、Schmidt et al. (2016a)と同様に、結果を 5 年ごとに集計した。ただし、21～25 歳の回答者はごく僅かであったため、21～30 歳は合算して 10 年分とした。

表 3-34 に年齢層別データ公開経験の集計結果を示す。年齢層別にデータ公開経験がある回答者の比率（以下、「公開率」）を確認すると、比較的データ公開率が高いのは、66～70 歳（56.9%）、36～40 歳（55.9%）、56～60 歳（54.8%）であった。データ公開率が低いのは、21～30 歳（40.0%）、46～50 歳（42.0%）、61～65 歳（42.3%）であり、年齢層による偏りや傾向（年齢層が高いほど公開率が高いなど）はみられなかった。

表 3-34 年齢層別データ公開経験

年齢	あり	なし	わからない	使わない	合計
21-30 歳	12 40.0%	17 56.7%	1 3.3%	0 0.0%	30 100.0%
31-35 歳	94 48.7%	97 50.3%	1 0.5%	1 0.5%	193 100.0%
36-40 歳	254 55.9%	192 42.3%	8 1.8%	0 0.0%	454 100.0%
41-45 歳	133 51.0%	123 47.1%	5 1.9%	0 0.0%	261 100.0%
46-50 歳	63 42.0%	82 54.7%	5 3.3%	0 0.0%	150 100.0%
51-55 歳	47 49.5%	47 49.5%	1 1.1%	0 0.0%	95 100.0%
56-60 歳	51 54.8%	41 44.1%	0 0.0%	1 1.1%	93 100.0%
61-65 歳	30 42.3%	37 52.1%	4 5.6%	0 0.0%	71 100.0%
66-70 歳	29 56.9%	20 39.2%	2 3.9%	0 0.0%	51 100.0%
合計	713 51.0%	656 46.9%	27 1.9%	2 0.1%	1,398 100.0%

年齢による差がみられた項目は、データの提供・被提供経験、カレントデータのプロフィール（データ保存期間規定の有無、データの量、データの機密情報の有無）などであった。差がみられた9項目は、まとめて表 3-61 に示す。ここでは、全体で最も懸念が強かった「先に論文を出版される可能性」と「引用せずに利用される可能性」、およびデータの提供・被提供経験の結果を示す。

図 3-27 に年齢層別の「先に論文を出版される可能性」への懸念の強さを示す。排列は年齢層が低い順とした（以下、同様に示す）。「問題である」と「やや問題である」の選択率の合計は、おおむね年齢層が上がるにつれて低くなることが明らかになった（51～55 歳の合計 79.0%と、56～60 歳の合計 80.4%のみ逆転していた）。「問題である」の選択率が最も高いのは 31～35 歳（70.3%）であり、次いで 36～40 歳（68.1%）が高かった。

3 章 日本の研究者によるデータ公開状況と内的要因

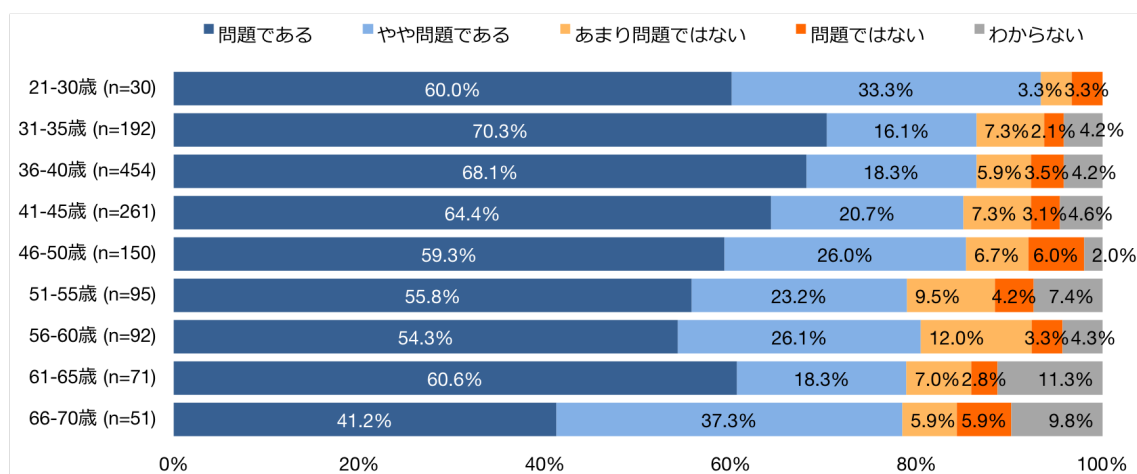


図 3-27 年齢層別「先に論文を出版される可能性」への懸念 (n=1,396)

図 3-28 に年齢層別の「引用せずに利用される可能性」への懸念の強さを示す。「先に論文を出版される可能性」ほどではないが、「問題である」と「やや問題である」の選択率の合計は、年齢層が上がるにつれて低くなる傾向がみられた（36～40 歳，61～65 歳などに逆転がみられる）。「問題である」の選択率が最も高いのは 36～40 歳（68.7%）であり，21～30 歳では「問題ではない」，「あまり問題ではない」は全く選ばれていなかった。

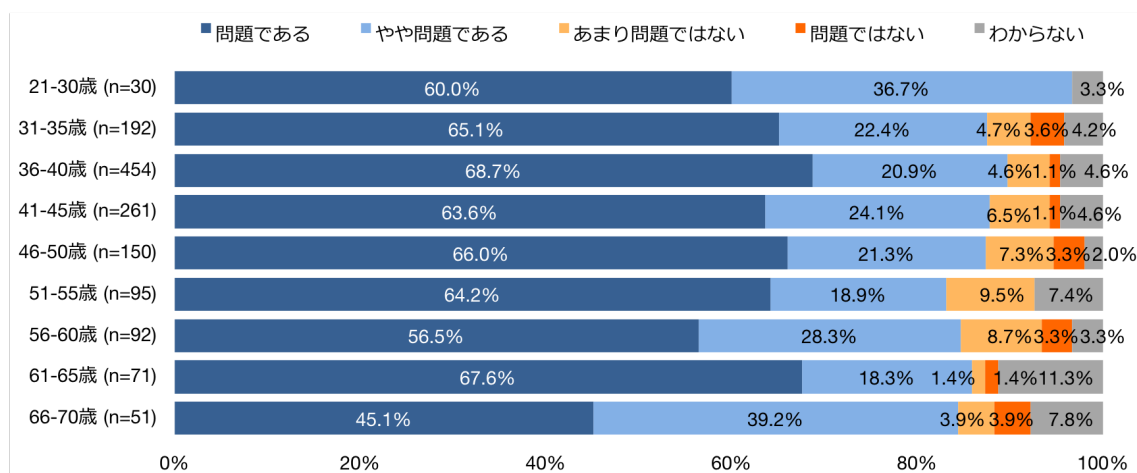


図 3-28 年齢層別「引用せずに利用される可能性」への懸念 (n=1,396)

図 3-29 に，年齢層別のデータ提供経験の頻度を示す。おおむね年齢層が高くなるほど提供頻度が高くなる傾向があることがわかった。66～70 歳は「よくある」の選択率（13.7%）も，「よくある」と「たまにある」の合計（56.8%）も最も高かった。

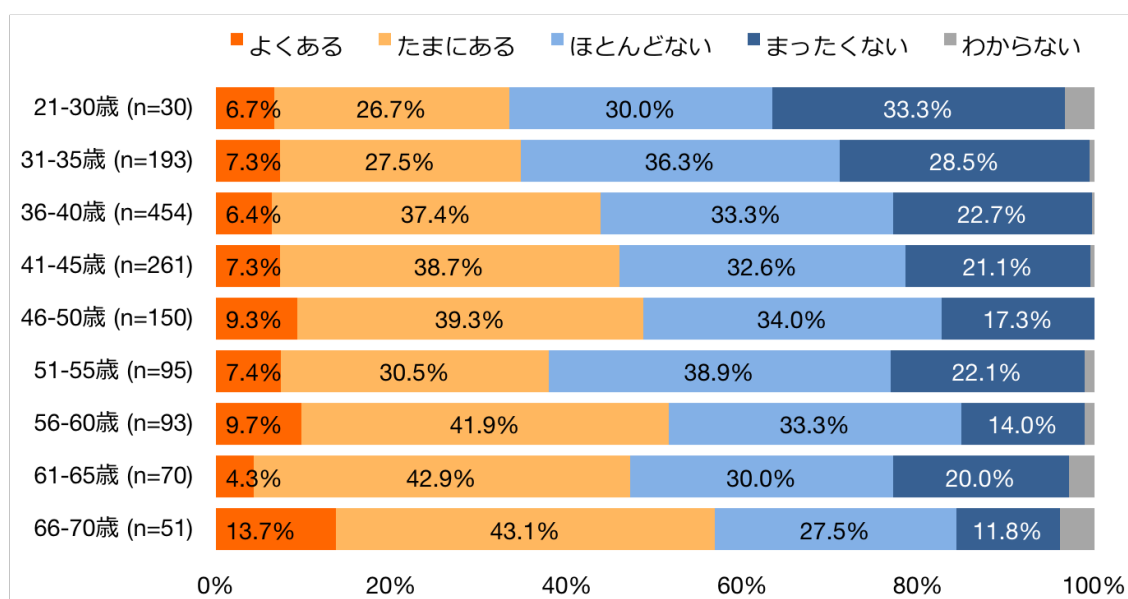


図 3-29 年齢層別データ提供経験 (n=1,397)

図 3-30 に、年齢層別のデータ提供経験の頻度を示す。提供経験 (図 3-29) と比較すると、年齢層の高さと被提供経験の頻度の高さの関連はゆるやかであり、21～30 歳の研究者は「よくある」の選択率が最も高かった (10.0%)。

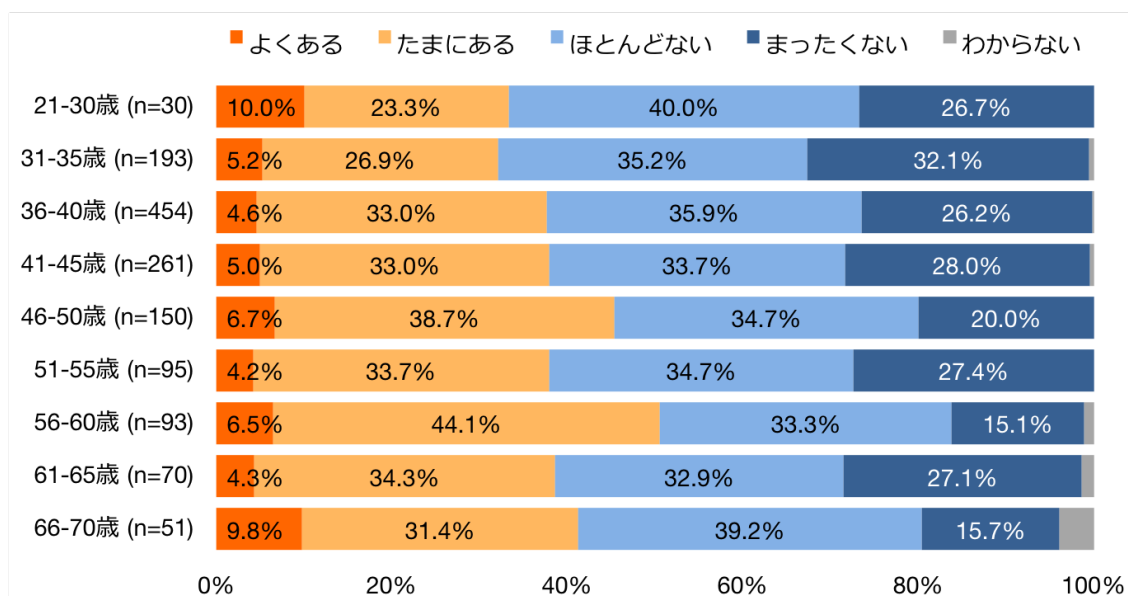


図 3-30 年齢層別データ被提供経験 (n=1,397)

3.2.10 データ公開経験の有無による比較

3.2.4 から 3.2.7 までの項目について、データ公開経験の有無ごとに集計して、差があるかどうかを確認した。また、次節（3.3 考察）の表 3-61 に差がみられた項目の一覧を示す。

[1] データの整備・公開のための人材の充足度とデータ公開経験の有無

表 3-35 にデータの整備・公開のための人材の充足度とデータ公開経験の有無のクロス集計結果を示す。人材が「充分」であると認識している回答者のうち、データ公開経験がある回答者は 85.7%，ない回答者は 14.3%であった。「ほぼ充分」は 78.3%，「やや不充分」は 60.7%，「不充分」は 49.6%であり、充足度が高いと認識している回答者ほど公開率が高かった。

表 3-35 データの整備・公開のための人材の充足度とデータ公開経験

充足度	公開		あり		なし		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
充分	12	85.7%	2	14.3%	14	100.0%		
ほぼ充分	47	78.3%	13	21.7%	60	100.0%		
やや不充分	108	60.7%	70	39.3%	178	100.0%		
不充分	451	49.6%	458	50.4%	909	100.0%		
わからない	95	45.7%	113	54.3%	208	100.0%		
合計	713	52.1%	656	47.9%	1,369	100.0%		

[2] 時間の充足度とデータ公開経験の有無

表 3-36 に時間の充足度とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、時間が「充分」だと認識している回答者は 81.0%，「ほぼ充分」は 78.7%，「やや不充分」は 62.2%，「不充分」は 46.2%であり、充足度が高いと認識している回答者ほど公開率が高かった。

表 3-36 データの整備・公開のための時間の充足度とデータ公開経験

公開 充足度	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
充分	17	81.0%	4	19.0%	21	100.0%
ほぼ充分	74	78.7%	20	21.3%	94	100.0%
やや不充分	158	62.2%	96	37.8%	254	100.0%
不充分	355	46.2%	413	53.8%	768	100.0%
わからない	109	47.0%	123	53.0%	232	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[3] 資金の充足度とデータ公開経験の有無

表 3-37 に資金の充足度とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、資金が「充分」だと認識している回答者は 72.7%、「ほぼ充分」は 68.4%、「やや不充分」は 64.8%、「不充分」は 50.1%であり、充足度が高いと認識している回答者ほど公開率が高かった。

表 3-37 データの整備・公開のための資金の充足度とデータ公開経験

公開 充足度	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
充分	16	72.7%	6	27.3%	22	100.0%
ほぼ充分	52	68.4%	24	31.6%	76	100.0%
やや不充分	129	64.8%	70	35.2%	199	100.0%
不充分	410	50.1%	409	49.9%	819	100.0%
わからない	106	41.9%	147	58.1%	253	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[4] データ保存用ストレージの充足度とデータ公開経験の有無

表 3-38 にデータ保存用ストレージの充足度とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、データ保存用ストレージが「充分」だと認識している回答者は 56.5%、「ほぼ充分」は 51.1%、「やや不充分」は 54.9%、「不充分」は 51.5%であり、

3 章 日本の研究者によるデータ公開状況と内的要因

充足度が高いと認識している回答者ほど公開率が高いとはいえなかった。

表 3-38 データ保存用ストレージの充足度とデータ公開経験

充足度 \ 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
充分	61	56.5%	47	43.5%	108	100.0%
ほぼ充分	144	51.1%	138	48.9%	282	100.0%
やや不充分	150	54.9%	123	45.1%	273	100.0%
不充分	264	51.5%	249	48.5%	513	100.0%
わからない	94	48.7%	99	51.3%	193	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[5] データ公開用リポジトリの充足度とデータ公開経験の有無

表 3-39 にデータ公開用リポジトリの充足度とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、データ公開用リポジトリが「充分」だと認識している回答者は 80.3%, 「ほぼ充分」は 63.2%, 「やや不充分」は 61.0%, 「不充分」は 46.6%であり、充足度が高いと認識している回答者ほど公開率が高かった。

表 3-39 データ公開用リポジトリの充足度とデータ公開経験

充足度 \ 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
充分	49	80.3%	12	19.7%	61	100.0%
ほぼ充分	120	63.2%	70	36.8%	190	100.0%
やや不充分	136	61.0%	87	39.0%	223	100.0%
不充分	245	46.6%	281	53.4%	526	100.0%
わからない	163	44.2%	206	55.8%	369	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[6] 研究中のデータ用ストレージの充足度とデータ公開経験の有無

表 3-40 に研究中のデータ用ストレージの充足度とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、研究中のデータ用ストレージが「充分」だと認識

している回答者は 56.6%, 「ほぼ充分」は 51.5%, 「やや不充分」は 57.2%, 「不充分」は 49.8% であり, 充足度が高いと認識している回答者ほど公開率が高いとはいえなかった。

表 3-40 研究中のデータ用ストレージの充足度とデータ公開経験

充足度 \ 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
充分	82	56.6%	63	43.4%	145	100.0%
ほぼ充分	195	51.5%	184	48.5%	379	100.0%
やや不充分	135	57.2%	101	42.8%	236	100.0%
不充分	209	49.8%	211	50.2%	420	100.0%
わからない	92	48.7%	97	51.3%	189	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[7] 引用せずに利用される可能性とデータ公開経験の有無

表 3-41 に引用せずに利用される可能性とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は, 引用せずに利用される可能性が「問題ではない」と認識している回答者は 65.4%, 「あまり問題ではない」は 63.2%, 「やや問題である」は 52.5%, 「問題である」は 51.2%であった。つまり, 懸念が弱い回答者ほど公開率が高かった。

表 3-41 引用せずに利用される可能性とデータ公開経験

懸念 \ 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
問題ではない	17	65.4%	9	34.6%	26	100.0%
あまり問題ではない	48	63.2%	28	36.8%	76	100.0%
やや問題である	166	52.5%	150	47.5%	316	100.0%
問題である	454	51.2%	433	48.8%	887	100.0%
わからない	28	43.8%	36	56.3%	64	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[8] 先に論文を出版される可能性とデータ公開経験の有無

表 3-42 に先に論文を出版される可能性とデータ公開経験の有無のクロス集計結果を示す。

3 章 日本の研究者によるデータ公開状況と内的要因

データ公開経験をもつ比率は、先に論文を出版される可能性が「問題ではない」と認識している回答者は 63.3%、「あまり問題ではない」は 63.3%、「やや問題である」は 51.4%、「問題である」は 51.0%であった。つまり、懸念が弱い回答者ほどやや公開率が高い程度であった。

表 3-42 先に論文を出版される可能性とデータ公開経験

懸念	公開		あり		なし		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
問題ではない	31	63.3%	18	36.7%	49	100.0%		
あまり問題ではない	62	63.3%	36	36.7%	98	100.0%		
やや問題である	148	51.4%	140	48.6%	288	100.0%		
問題である	444	51.0%	426	49.0%	870	100.0%		
わからない	28	43.8%	36	56.3%	64	100.0%		
合計	713	52.1%	656	47.9%	1,369	100.0%		

[9] 誤解や誤用の可能性とデータ公開経験の有無

表 3-43 に誤解や誤用の可能性とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、誤解や誤用の可能性が「問題ではない」と認識している回答者は 55.1%、「あまり問題ではない」は 50.8%、「やや問題である」は 55.7%、「問題である」は 49.6%であり、懸念が弱い回答者ほど公開率が高いとはいえなかった。

表 3-43 誤解や誤用の可能性とデータ公開経験

懸念	公開		あり		なし		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
問題ではない	54	55.1%	44	44.9%	98	100.0%		
あまり問題ではない	91	50.8%	88	49.2%	179	100.0%		
やや問題である	246	55.7%	196	44.3%	442	100.0%		
問題である	283	49.6%	287	50.4%	570	100.0%		
わからない	39	48.8%	41	51.2%	80	100.0%		
合計	713	52.1%	656	47.9%	1,369	100.0%		

[10] 機密・プライバシー情報の懸念とデータ公開経験の有無

表 3-44 に機密・プライバシー情報の懸念とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、機密・プライバシー情報が「問題ではない」と認識している回答者は 56.5%、「あまり問題ではない」は 58.1%、「やや問題である」は 54.5%、「問題である」は 48.8%であった。懸念が弱いほど公開率が高い傾向にあるが、「あまり問題ではない」と認識している回答者の公開率が最も高かった。

表 3-44 機密・プライバシー情報の懸念とデータ公開経験

懸念 \ 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
問題ではない	87	56.5%	67	43.5%	154	100.0%
あまり問題ではない	86	58.1%	62	41.9%	148	100.0%
やや問題である	164	54.5%	137	45.5%	301	100.0%
問題である	333	48.8%	350	51.2%	683	100.0%
わからない	43	51.8%	40	48.2%	83	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[11] 商用利用される可能性とデータ公開経験の有無

表 3-45 に商用利用される可能性とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、商用利用される可能性が「問題ではない」と認識している回答者は 55.8%、「あまり問題ではない」は 61.2%、「やや問題である」は 55.4%、「問題である」は 48.4%であった。懸念が弱いほど公開率が高い傾向にあるが、機密・プライバシー情報と同様に、「あまり問題ではない」と認識している回答者の公開率が最も高かった。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-45 商用利用される可能性とデータ公開経験

懸念	公開		あり		なし		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
問題ではない	82	55.8%	65	44.2%	147	100.0%		
あまり問題ではない	101	61.2%	64	38.8%	165	100.0%		
やや問題である	181	55.4%	146	44.6%	327	100.0%		
問題である	304	48.4%	324	51.6%	628	100.0%		
わからない	45	44.1%	57	55.9%	102	100.0%		
合計	713	52.1%	656	47.9%	1,369	100.0%		

[12] 研究の誤りを発見される可能性とデータ公開経験の有無

表 3-46 に研究の誤りを発見される可能性とデータ公開経験の有無のクロス集計結果を示す。データ公開経験をもつ比率は、研究の誤りを発見される可能性が「問題ではない」と認識している回答者は 54.0%、「あまり問題ではない」は 51.9%、「やや問題である」は 49.2%、「問題である」は 50.0%であった。「問題ではない」は、やや公開率が高かったものの、その他はほとんど差がなかった。

表 3-46 研究の誤りを発見される可能性とデータ公開経験

懸念	公開		あり		なし		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
問題ではない	281	54.0%	239	46.0%	520	100.0%		
あまり問題ではない	294	51.9%	272	48.1%	566	100.0%		
やや問題である	58	49.2%	60	50.8%	118	100.0%		
問題である	39	50.0%	39	50.0%	78	100.0%		
わからない	41	47.1%	46	52.9%	87	100.0%		
合計	713	52.1%	656	47.9%	1,369	100.0%		

[13] データ提供の頻度とデータ公開経験の有無

表 3-47 にデータ提供の頻度とデータ公開経験のクロス集計結果を示す。データ提供経験が「よくある」回答者のうち、データ公開経験をもつ回答者は 75.5%、「たまにある」回答者は

65.4%,「ほとんどない」回答者は 45.1%,「まったくない」回答者は 32.1%であった。つまり、データ提供の頻度が高い回答者ほど、データ公開経験をもつ比率が高かった。

表 3-47 データ提供経験とデータ公開経験

提供 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
よくある	77	75.5%	25	24.5%	102	100.0%
たまにある	329	65.4%	174	34.6%	503	100.0%
ほとんどない	206	45.1%	251	54.9%	457	100.0%
まったくない	97	32.1%	205	67.9%	302	100.0%
合計	709	52.0%	655	48.0%	1,364	100.0%

[14] データの被提供の頻度とデータ公開経験の有無

表 3-48 にデータの被提供の頻度とデータ公開経験の有無のクロス集計結果を示す。データ提供経験が「よくある」回答者のうち、データ公開経験をもつ回答者は 67.6%,「たまにある」回答者は 65.6%,「ほとんどない」回答者は 48.7%,「まったくない」回答者は 35.7%であった。すなわち、データの被提供経験も、頻度が高い回答者ほどデータ公開経験をもつ比率が高かった。

表 3-48 データ被提供経験とデータ公開経験

被提供 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
よくある	50	67.6%	24	32.4%	74	100.0%
たまにある	300	65.6%	157	34.4%	457	100.0%
ほとんどない	233	48.7%	245	51.3%	478	100.0%
まったくない	127	35.7%	229	64.3%	356	100.0%
合計	710	52.0%	655	48.0%	1,365	100.0%

[15] 公開データの入手経験の有無とデータ公開経験の有無

表 3-49 に公開データの入手経験の有無とデータ公開経験の有無のクロス集計結果を示す。公開データの入手経験をもつ回答者のうち、データ公開経験をもつ回答者は 59.9%, 入手経

3 章 日本の研究者によるデータ公開状況と内的要因

験がない回答者のうちデータ公開経験がない回答者は 73.6%であった。

表 3-49 公開データの入手経験とデータ公開経験

入手 公開	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
あり	626	59.9%	419	40.1%	1,045	100.0%
なし	82	26.4%	229	73.6%	311	100.0%
合計	708	52.2%	648	47.8%	1,356	100.0%

[16] 公開データの利用状況とデータ公開経験の有無

表 3-50 に公開データの利用状況とデータ公開経験の有無のクロス集計表を示す。「研究の参考」、「研究に再利用」、「研究を再現」のいずれも、行う頻度が高い回答者の方がデータ公開経験をもつ回答者の比率が高いことが明らかになった。最も比率が高かったのは、公開データを「研究に再利用」する回答者であり、68.2%であった。

表 3-50 データの利用状況とデータ公開経験

		公開		あり		なし		合計	
		人数	比率	人数	比率	人数	比率		
利用	行う	250	66.5%	126	33.5%	376	100.0%		
	たまに行う	328	56.9%	248	43.1%	576	100.0%		
	行わない	48	51.6%	45	48.4%	93	100.0%		
研究の参考	行う	120	68.2%	56	31.8%	176	100.0%		
	たまに行う	245	60.9%	157	39.1%	402	100.0%		
	行わない	261	55.9%	206	44.1%	467	100.0%		
研究に再利用	行う	59	66.3%	30	33.7%	89	100.0%		
	たまに行う	248	62.9%	146	37.1%	394	100.0%		
	行わない	319	56.8%	243	43.2%	562	100.0%		
研究を再現	合計	626	59.9%	419	40.1%	1,045	100.0%		

[17] 公開データの利用希望の有無とデータ公開経験の有無

表 3-51 に公開データの利用希望の有無とデータ公開経験の有無のクロス集計結果を示す。

公開データの利用を希望する回答者（分野を1つ以上選択した回答者）のうち、データ公開経験をもつ回答者は52.9%、もたない回答者は47.1%であった。

表 3-51 公開データの利用希望とデータ公開経験

公開 利用希望	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
あり	702	52.9%	626	47.1%	1,328	100.0%
なし	11	26.8%	30	73.2%	41	100.0%
合計	713	52.1%	656	47.9%	1,369	100.0%

[18] カレントデータの量とデータ公開経験の有無

表 3-52 にカレントデータの量とデータ公開経験の有無のクロス集計結果を示す。「MB 以下」は51.4%、「GB」は51.3%、「TB」は54.8%、「PB 以上」は75.0%（ただし回答者は4名）であり、データ量が大きい回答者ほど、わずかにデータ公開経験をもつ比率が高かった。

表 3-52 カレントデータの量とデータ公開経験の有無

公開 データ量	ある		ない		合計	
	人数	比率	人数	比率	人数	比率
MB 以下	133	51.4%	126	48.6%	259	100.0%
GB	394	51.3%	374	48.7%	768	100.0%
TB	103	54.8%	85	45.2%	188	100.0%
PB 以上	3	75.0%	1	25.0%	4	100.0%
合計	633	51.9%	586	48.1%	1,219	100.0%

[19] カレントデータの所有権とデータ公開経験の有無

表 3-53 に、自分以外にカレントデータの所有権をもつ人や組織の有無とデータ公開経験の有無のクロス集計結果を示す。自分以外に所有権を持つ人や組織がある回答者の方が、わずかにデータ公開経験をもつ比率が高かった。

3 章 日本の研究者によるデータ公開状況と内的要因

表 3-53 カレントデータの所有権とデータ公開経験の有無

公開 所有権	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
あり	612	53.2%	538	46.8%	1,150	100.0%
なし	57	50.9%	55	49.1%	112	100.0%
合計	669	53.0%	593	47.0%	1,262	100.0%

[20] カレントデータの機密情報とデータ公開経験の有無

表 3-54 に、カレントデータに機密情報が含まれているかどうかとデータ公開経験の有無のクロス集計結果を示す。データ公開経験がある回答者については、機密情報がある回答者よりも機密情報がない回答者の比率の方がわずかに高かった。

表 3-54 カレントデータの機密情報とデータ公開経験

公開 機密情報	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
あり	340	50.5%	333	49.5%	673	100.0%
なし	261	53.5%	227	46.5%	488	100.0%
合計	601	51.8%	560	48.2%	1,161	100.0%

[21] カレントデータの理解の難しさとデータ公開経験の有無

表 3-55 に、同じ分野の研究者によるカレントデータの理解の難しさとデータ公開経験の有無のクロス集計表を示す。データ公開経験をもつ回答者の比率は、「難しいと思う」と考えている回答者は 43.3%、「やや難しいと思う」は 52.5%、「できると思う」は 54.1%であり、理解が容易であると考えている回答者の方がデータ公開経験の比率がわずかに高かった。

表 3-55 同じ分野の研究者によるカレントデータの理解とデータ公開経験

理解	公開		あり		なし		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
難しいと思う	61	43.3%	80	56.7%	141	100.0%		
やや難しいと思う	168	52.5%	152	47.5%	320	100.0%		
できると思う	447	54.1%	379	45.9%	826	100.0%		
合計	676	52.5%	611	47.5%	1,287	100.0%		

表 3-56 に、異分野の研究者によるカレントデータの理解の難しさとデータ公開経験の有無のクロス集計表を示す。データ公開経験をもつ回答者の比率は、「難しいと思う」と考えている回答者は 48.0%、「やや難しいと思う」は 58.3%、「できると思う」は 51.7%であり、「やや難しいと思う」と考えている回答者のデータ公開経験の比率が最も高かった。

表 3-56 異分野の研究者によるカレントデータの理解とデータ公開経験

理解	公開		あり		なし		合計	
	人数	比率	人数	比率	人数	比率	人数	比率
難しいと思う	282	48.0%	306	52.0%	588	100.0%		
やや難しいと思う	299	58.3%	214	41.7%	513	100.0%		
できると思う	93	51.7%	87	48.3%	180	100.0%		
合計	674	52.6%	607	47.4%	1,281	100.0%		

[1]から[21]までの結果により、データの公開経験の有無と関連があると考えられる 15 項目が明らかになった。このうち、どのような要因の組み合わせをもつ研究者がデータを公開しているのかを明らかにするために、経験の有無を目的変数として 2 項ロジスティック回帰分析を行った。

まず、15 項目から説明変数の候補を選択した（表 3-57）。多重共線性が生じないよう項目同士の相関分析を行い¹⁸⁾、0.4 以上の相関がみられる項目はデータ公開経験との関連がより高い方の項目を残して除外した（5 項目）。公開データの利用目的は、公開データの入手経験がある回答者にのみ尋ねているため、片方ずつ回帰式に投入して、あてはまりの良い方を採用することとした。また、線形回帰分析によって VIF 値が高い項目がないことを確認した。完全分離の状態にある項目もなかった。

¹⁸⁾ ①OA 経験と②公開データの入手経験のみ Pearson の相関係数を、その他の組み合わせは全て Spearman の相関係数を用いた。

表 3-57 データ公開経験の有無と関連のある項目

説明変数（候補）
① OA 経験
② 公開データの入手経験
②' 公開データの利用目的：②'-1 研究の参考，②'-2 研究に再利用，②'-3 研究を再現
③ データ提供経験
④ データを整備・公開するための時間
⑤ 商用利用される可能性
⑥ 異分野の研究者による理解
⑦ 所属（3=大学，2=公的機関・団体，1=企業）※データ公開経験ありの回答率が高い順序
除外した項目
・ データ被提供経験（③と相関あり $r=0.77$, $p<0.01$ ）
・ データ公開用のリポジトリ（④と相関あり $r=0.49$, $p<0.01$ ）
・ データを整備・公開するための人材（④と相関あり $r=0.71$, $p<0.01$ ）
・ データを整備・公開するための資金（④と相関あり $r=0.56$, $p<0.01$ ）
・ 機密・プライバシー情報（⑤と相関あり $r=0.60$, $p<0.01$ ）

こうして得られた7項目（②を含む）と9項目（②'を含む）を説明変数として，変数増加法による分析を行った。順序尺度は最小の値（「まったくない」などネガティブな選択肢）を参照カテゴリとした。ステップワイズにおける有意確率を，投入=0.05，除去=0.10 として実施した結果，「OA 経験」，「公開データの入手経験」，「データ提供経験」，「データを整備・公開するための時間」，「所属」の5変数による回帰式が得られた。しかし，各変数の欠損値によって全体の欠損値が大きくなるため，改めて5変数を用いた分析を行った。表 3-58 に5変数を用いた結果を示す（データ公開経験「あり」=577，「なし」=498）。

表 3-58 データ公開に影響を与える要因

	偏回帰係数	Wald χ^2	有意確率
OA 経験（あり）	0.605	13.422	0.000
公開データの入手経験（あり）	1.227	46.404	0.000
データ提供経験（まったくない）		58.696	0.000
データ提供経験（ほとんどない）	0.431	4.767	0.029
データ提供経験（たまにある）	1.258	41.351	0.000
データ提供経験（よくある）	1.572	26.381	0.000
時間（不十分）		54.572	0.000
時間（やや不十分）	0.873	24.685	0.000
時間（ほぼ充分）	1.710	33.524	0.000
時間（充分）	1.494	6.341	0.012
所属（企業）		27.762	0.000
所属（公的機関・団体）	0.824	12.23	0.000
所属（大学）	0.998	27.65	0.000
定数	-3.136	118.644	0.000

以上の結果から、データ公開経験をもつ研究者の特徴は、OA 経験、公開データの入手経験、データの提供経験があり、データを整備・公開するための時間が比較的充分であり、大学や公的機関・団体に所属していると予測できる。回帰式によるデータ公開経験の有無の予測正解率は、あり=75.6%、なし=64.1%、全体=70.2%であった。

3.2.11 データ公開意思の有無による比較

Q21 では、データ公開経験がなく、かつ、何らかの非公開理由をもつ回答者（595 名）に「非公開理由」が解決した場合のデータ公開意思を尋ねた（3.2.3）。3.2.4 から 3.2.7 までの項目のうち、データ公開意思の有無¹⁹⁾、すなわちデータの公開意思が「あり」の回答者（169 名）と「なし」の回答者（130 名）で差がみられた項目は、データ公開への懸念、個人的なデータの提供経験、データの被提供経験、公開データの入手経験、公開データの再分析・再利用経験、公開データの利用希望であった。順に結果を示す。なお、結果のまとめは表 3-61 に示す。

¹⁹⁾ 「わからない」を選択した 296 名は除外した。

[1] データ公開への懸念

データ公開意思の有無によるデータ公開への懸念は、「引用せずに利用される可能性」、「先に論文を出版される可能性」、「誤解や誤用の可能性」、「商用利用される可能性」に差がみられた。一方、データ公開意思の有無による「機密・プライバシー情報」、「研究の誤りを発見される可能性」、およびデータを公開する場合の資源の充足度には差がみられなかった。

図 3-31 に、データの公開意思の有無と「引用せずに利用される可能性」への懸念の強さを示す。「引用せずに利用される可能性」を「問題である」とする回答者は、データ公開意思がある回答者は 62.1%、ない回答者は 76.2%であった。「やや問題である」とする回答者は、データ公開意思がある回答者は 25.4%、ない回答者は 10.8%であった。

差がみられた項目のうち、この「引用せずに利用される可能性」のみ、公開意思をもつ回答者が「問題である」と「やや問題である」を選択した比率の合計が高かった。すなわち、データ公開意思がある回答者は 87.5%、ない回答者は 87.0%であった。

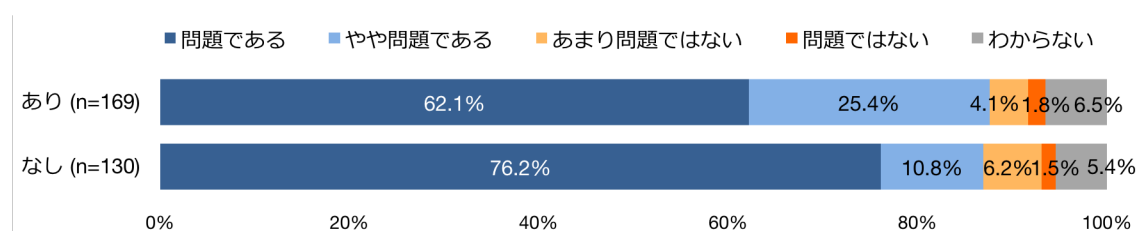


図 3-31 データ公開意思の有無と「引用せずに利用される可能性」への懸念

図 3-32 に、データの公開意思の有無と「先に論文を出版される可能性」への懸念の強さを示す。「先に論文を出版される可能性」を「問題である」とする回答者は、データ公開意思がある回答者は 56.8%、ない回答者は 75.4%であった。「やや問題である」とする回答者は、データ公開意思がある回答者は 26.0%、ない回答者は 13.8%であった。

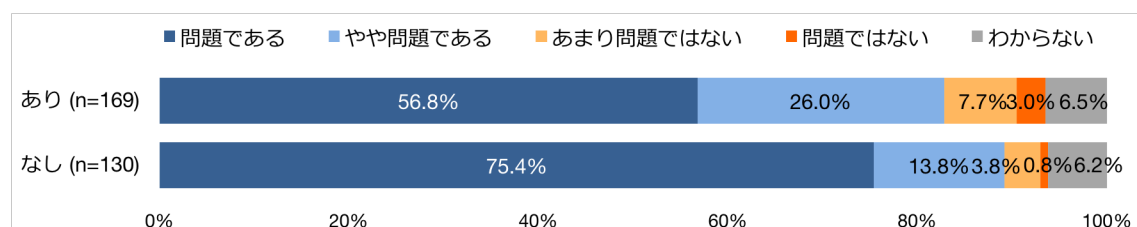


図 3-32 データ公開意思の有無と「先に論文を出版される可能性」への懸念

図 3-33 に、データの公開意思の有無と「誤解や誤用の可能性」への懸念の強さを示す。「誤解や誤用の可能性」を「問題である」とする回答者は、データ公開意思がある回答者は 36.1%、

ない回答者は 57.7%であった。「やや問題である」とする回答者は、データ公開意思がある回答者は 29.6%，ない回答者は 22.3%であった。

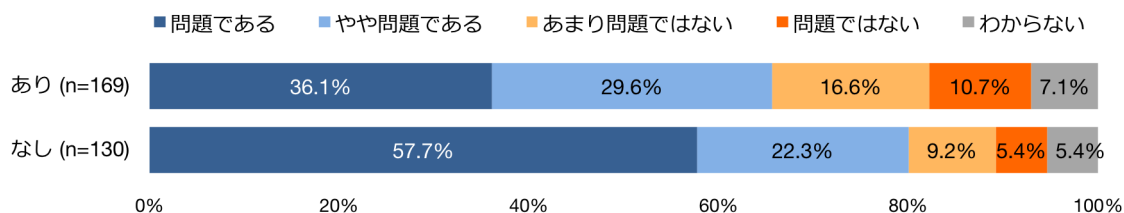


図 3-33 データ公開意思の有無と「誤解や誤用の可能性」への懸念

図 3-34 に、データの公開意思の有無と「商用利用される可能性」への懸念の強さを示す。「商用利用される可能性」を「問題である」とする回答者は、データ公開意思がある回答者は 41.4%，ない回答者は 61.5%であった。「やや問題である」とする回答者は、データ公開意思がある回答者は 24.3%，ない回答者は 15.4%であった。

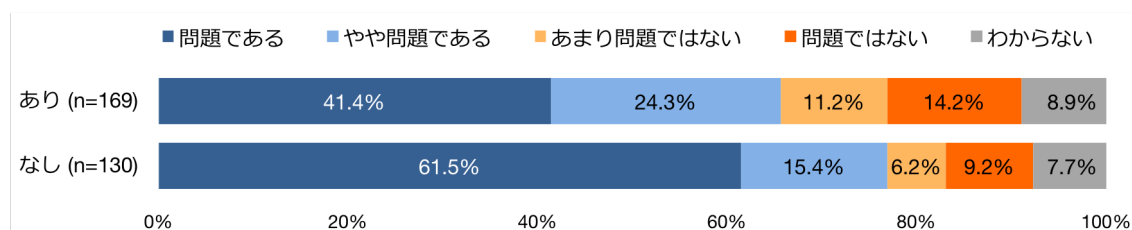


図 3-34 データ公開意思の有無と「商用利用される可能性」への懸念

[2] データの提供・被提供経験

図 3-35 に、データ公開意思の有無とデータの提供経験を示す。データの提供経験が「よくある」を選んだ回答者の比率（以下、「選択率」）は、データ公開意思がある回答者は 4.7%，ない回答者は 3.1%であった。「たまにある」の選択率は、データ公開意思がある回答者は 29.0%，ない回答者は 16.2%であり、いずれもデータ公開意思がある回答者の比率が高かった。

3 章 日本の研究者によるデータ公開状況と内的要因

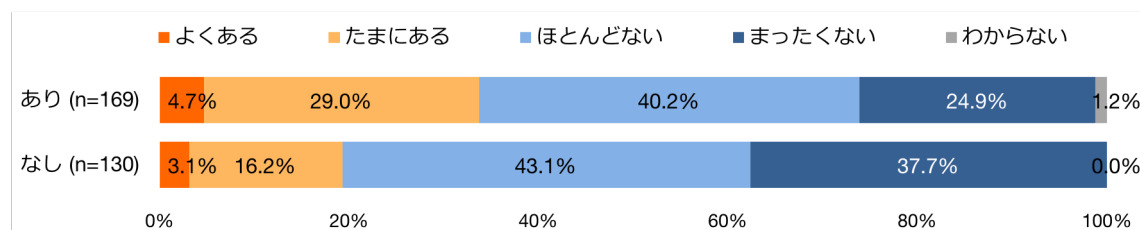


図 3-35 データ公開意思の有無とデータ提供経験

図 3-36 に、データの公開意思の有無とデータの被提供経験を示す。データの被提供経験が「よくある」の選択率は、データ公開意思がある回答者は 4.1%，ない回答者は 3.1%であった。「たまにある」の選択率は、データ公開意思がある回答者は 32.0%，ない回答者は 12.3%であり、いずれもデータ公開意思がある回答者の比率が高かった。

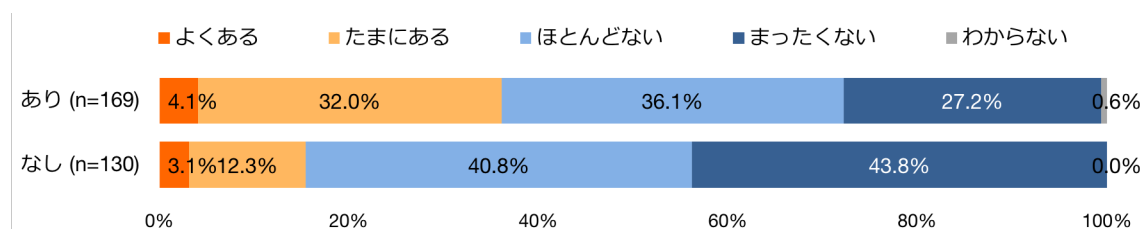


図 3-36 データ公開意思の有無とデータ被提供経験

[3] 公開データの入手経験

表 3-59 に、データの公開意思の有無と公開データの入手経験の有無の集計結果を示す。データの入手経験は、データ公開意思がある回答者の方が多く（68.3%），ない回答者は 52.7%であった。

表 3-59 データ公開意思の有無とデータ入手経験の有無

データ入手 公開意思	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
あり	114	68.3%	53	31.7%	167	100.0%
なし	68	52.7%	61	47.3%	129	100.0%
合計	182	61.5%	114	38.5%	296	100.0%

[4] 公開データの再分析・再利用経験

図 3-37 に、データ公開意思の有無と公開データの再分析・再利用経験を示す。データの再分析・再利用を「行う」とする回答者は、データ公開意思がある回答者は 19.3%，ない回答者は 8.8%であった。「たまに行う」とする回答者は、データ公開意思がある回答者は 45.6%，ない回答者は 27.9%であり、いずれもデータ公開意思がある回答者の比率が高かった。

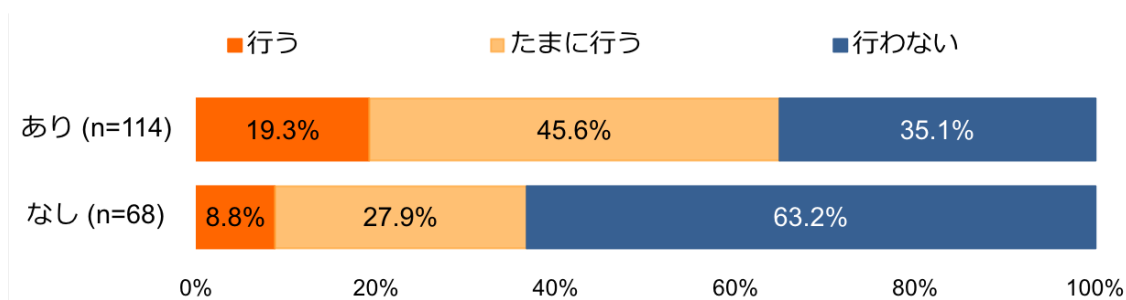


図 3-37 データ公開意思の有無と公開データの再分析・再利用経験

[5] 公開データの利用希望

表 3-60 に、データ公開意思の有無と公開データの利用希望の有無のクロス集計表を示す。データ公開意思をもつ回答者のうち、公開データの利用を希望する回答者（分野を 1 つ以上選択した回答者）は 99.4%，もたない回答者は 89.2%であった。

表 3-60 データ公開意思の有無と公開データの利用希望の有無						
公開意思 \ 利用希望	あり		なし		合計	
	人数	比率	人数	比率	人数	比率
あり	168	99.4%	1	0.6%	169	100.0%
なし	116	89.2%	14	10.8%	130	100.0%
合計	284	95.0%	15	5.0%	299	100.0%

3.3 考察

ここでは、調査結果の考察を行う。質問紙調査全体については、回答率が高く（70.9%），

必須ではない質問であっても無回答の比率が低かった（0%～1.7%）。また、データや論文を公開した経験がない回答者からも回答を得ることができた。しかし、自由回答では「問題の量が多い」という指摘もみられた。

3.3.1 日本の研究者によるデータ公開の状況

[1] 日本の研究者のデータ公開率

本研究の分析対象とした 1,398 名のうち、データの公開経験をもつ回答者（以下、「データ公開率」）は 713 名であった（表 3-5）。すなわち、日本の研究者によるデータ公開率は 51.0% であり、大学の研究者に限ると 56.9% であった（表 3-31）。

に示した先行研究のうち、多分野の研究者を対象とした調査の結果と本調査の結果を比較する。2014 年に Wiley 社が世界の研究者 2,250 名以上から回答を得た調査によれば、データ公開率は 52% であり、日本の研究者²⁰⁾は 44% であった（Ferguson, 2014）。2016 年にライデン大学の科学技術研究センター（Centre for Science and Technology Studies, CWTS）と Elsevier 社が 1,162 名から回答を得た調査（以下、「CWTS/Elsevier 調査」）では、データ公開率は 66%²¹⁾ であった（Berghmans et al., 2017）。Wiley 社の調査結果、および本研究とほぼ同時期に実施された CSTS/Elsevier 調査の結果との比較から、日本の研究者によるデータ公開率は、国外の研究者と比較してやや低いと考えられる。

[2] 日本の研究者のデータ公開方法

データの公開方法として最も選択率が高かったのは、「個人や研究室のウェブサイト」（50.8%）、次いで「論文の補足資料」（47.0%）であった。オープンサイエンス政策などではパブリックリポジトリでの公開が推奨されているが、「所属機関のリポジトリ」は 34.2%、「特定分野のリポジトリ」は 16.4% にとどまっていた（表 3-6）。

2.4.1 に示した先行研究は、選択肢や尋ね方が異なるものの、Huang et al. (2012)、Ferguson (2014)、Berghmans et al. (2017) の調査では論文の補足資料が最も多く、その他の調査でもリポジトリは第 1 位とはなっていなかった。リポジトリによる公開が主流とはいえないことは、現状では国内外に共通の課題であると考えられる。

²⁰⁾ 日本を含め、国別の回答者数については記載がなかった。

²¹⁾ 本調査と比較するため、さまざまなデータ公開方法の選択肢を全く選択していない研究者の比率（34%）を全体から減じた比率（66%）をデータ公開率として示した。

[3] 日本の研究者のデータ公開理由

データの公開経験をもつ 713 名の回答者にその理由を尋ねたところ、1 位は「研究成果の認知向上」(58.5%)であった(表 3-8)。また、自由記述でも研究成果をアピールするために公開するという積極的な姿勢がみられた(表 3-9)。2 位は「雑誌のポリシー」(43.8%)であり、現状では研究者に公開を促す制度よりも誘因の方が、やや選択率が高かった。

Kim and Stanton (2015) の調査では、科学的利他性が主なデータ公開理由の 1 つであった。科学的利他性に関連する「科学研究や成果実装の推進」は、Schmidt et al. (2016a), Damvad (2014), 小野ら (2016) の調査でも主な公開理由となっていたが、本調査では 3 位 (26.1%) にとどまった。また、「他の研究者からのリクエストに応じて」は 4 位 (25.4%), 「オープンデータへの貢献」は 6 位 (10.5%) であった。それぞれの選択率は OA 論文よりも高かったものの、主要な理由とまでは言い難かった。

「所属機関のポリシー」は 5 位 (23.7%), 「助成機関のポリシー」は 9 位 (6.3%) であり、やはり主要な理由ではなかった。米国、英国、EU では助成機関による義務化、すなわちデータマネジメントプラン (Data Management Plan, DMP) 提出の義務化によってデータ公開が進んでいると指摘されており (Fearon, et al., 2013; Whyte & Sisu, 2014), 先行研究 (質問紙調査) においてもデータ公開の動機とされていた (Wallis, 2013; Fecher et al., 2015; Kim & Burns, 2015; Schmidt et al., 2016a)。2016 年の調査時点では日本の助成機関はデータ公開を義務化していなかったが、2017 年に科学技術振興機構 (2017) が、2018 年には新エネルギー・産業技術総合開発機構 (2018) や日本医療研究開発機構 (2018) が DMP の提出を求めるようになったことから、これらを契機としてデータ公開が進むのではないかと予想される。所属機関のポリシーについては、3.3.6 (データ公開と論文の OA の差異) で述べる。

3.3.2 日本の研究者のデータ公開の障壁

[1] 日本の研究者によるデータの非公開理由と公開意思

データの非公開理由について、対象となる 683 名のうち 3 割以上の回答者が選択した項目はなかった(表 3-10)。2 割以上の回答者が選択した項目は、雑誌や所属機関にポリシーがないから、時間がないから、ニーズがないと思う、業績にならないからなど多岐にわたった。また、これらが解決した場合にデータ公開を実践しようとする回答者は 28.4%にとどまった(表 3-11)。データ公開意思がある回答者とない回答者に特徴的な非公開理由があるかどうかを確認するために、それぞれ集計したが偏りはみられなかった。つまり、これを解決すればデータ公開に繋がるという理由を特定することはできなかった。

自由記述では、公開のインセンティブの欠如が指摘された (8 件)。一方で、業績化や評価 (5 件)、引用 (3 件) が保証されるならばデータ公開に前向きになるとの意見もみられた。

先行研究ではデータ公開の業績化がデータ公開のインセンティブに繋がることが指摘され（Fecher et al., 2015; Ali-Khan, Harris, & Gold, 2017）、オープンサイエンス政策においても仕組みづくりが検討されている（Expert Advisory Group on Data Access, Economic and Social Research Council, Medical Research Council, Cancer Research UK, & Wellcome Trust, 2014; OECD, 2015; 国際的動向を踏まえたオープンサイエンスに関する検討会, 2015; 日本学術会議, 2016; G7 Science Ministers, 2017）。また、引用についても普及の取り組みが進められている（Starr et al., 2015）。この点は5章で改めて検討する。

非公開理由が解決された場合に公開意思がある回答者は 28.4%、ない回答者は 21.8%、わからないとした回答者は 49.7%であった。ここで挙げた非公開理由の他に本質的な非公開理由がある可能性、データ公開による利点やインセンティブが明確でない可能性などが考えられる。3.3.9 では、データの非公開理由が解決した場合であっても、データ公開を行わないとする回答者（データ公開に消極的な研究者）に着目して、その特徴について考察する。

[2] 日本の研究者によるデータ公開の障壁の程度

データ公開経験の有無にかかわらず、データ公開のための資源の充足度や懸念といった障壁が存在すると考えられるため、研究にデータを用いている回答者（1,396 名）に障壁の程度を尋ねた。データを整備し、公開しようとする場合の資源は、特に人材、資金、時間が不足していると認識されていた（図 3-1）。一方、データを公開している研究者は、公開のための時間の充足度が高い傾向にあった（表 3-36, 表 3-58）。

日本の研究者の時間が不足していることは、深刻な課題となっている（1.3.1 参照）。新たにデータ公開をすすめるのであれば、既に政策文書でも指摘されているように支援人材の育成が必要であると考えられる（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015; 文部科学省科学技術・学術審議会学術分科会第 8 期学術情報委員会, 2016; 日本学術会議, 2016; 内閣府, 2017）。1.4.3 でも触れたように、日本においても研究データ管理のための教材開発や DMP 作成支援ツールの開発（池内, 2018b）が進められている。本論文では触れなかったが、第三者が支援する際に「専門性が必要である」と考える回答者が少なかった機関リポジトリによる公開やメタデータの作成（付表 45）などから支援を開始することが考えられる。

資金については、データ公開の理由として記述されていたインフラの利用料金の割引（表 3-9, “データを公開すると実験設備の利用料が安くなる”）や、「内閣府報告書」に示されている“高品質なデータを提供した研究者に適切な報酬（助成金や昇進機会）を与える”（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015, p. 22）といった経済的支援の方法をとることができるならば、公開のインセンティブにも繋がると考えられる。

データを整備し、公開しようとする場合の懸念は、「研究の誤りを発見される可能性」以外、問題視する回答者が多かった（図 3-2）。特に「引用せずに利用される可能性」および「先に

論文を出版される可能性」は、データ公開経験の有無にかかわらず、最も問題視する回答者が多く、キャリアリスクを強く懸念していることがわかった。非公開理由では「業績にならないと思うから」の選択率が比較的高かったこと、業績、評価、引用の対象となるならばデータ公開をポジティブに捉えるという意見がみられたことから、データ公開を業績化することによって懸念を低減し、公開のインセンティブとすることができるのではないかと考えられる。ただし、“データを得ることは研究の一部ではあってもそれを得ただけで業績として扱われるのは、問題であると思う”という指摘もみられたことから、業績化や評価については熟議が必要であると考えられる。

3.3.3 データ公開と関連のある経験

[1] 日本の研究者によるデータの提供と被提供状況

全ての回答者（1,398 名）を対象に、データの提供経験と被提供経験の頻度、およびデータ公開経験の有無を調査したところ、いずれも頻度が高い回答者ほど、データ公開経験がある回答者の比率が高かった（3.2.5）。Damvad（2014）や Berghmans et al.（2017）はデータの提供経験がデータ公開に繋がると論じていた。本研究では両者の因果関係までは明らかにできなかったが、その可能性は示唆された。

[2] 日本の研究者による公開データの入手経験と問題

公開データの入手経験は、全回答者 1,398 名のうち 1,060 名（75.8%）が有していた（表 3-13）。公開データの入手先として選択率が高かったのは、「個人や研究室のウェブサイト」（64.8%）、「論文の補足資料」（53.1%）、「学術機関のリポジトリ・データアーカイブ」（50.0%）の順であり、データの公開先と一致していた。学術機関のリポジトリ・データアーカイブは、データの公開先としては 34.2%にとどまるものの、回答者の約半数が入手先として選択していたことから、ある程度認知されていると考えられる。また、表 3-49 に示した通り、公開データの利用経験がある回答者はデータ公開経験をもつ比率が高かった。

公開データの利用方法として、「研究の参考」、「研究に再利用」、「研究を再現」の頻度を調査した。その結果、Kratz and Strasser（2015b）と同様に「研究の参考」、「研究に再利用」、「研究を再現」の順に頻度が高かった（図 3-3）。また、Piwowar（2011）がデータの再利用経験とデータ公開経験に関連があることを示していたのと同様に、本研究においてもデータの再利用経験、すなわち「研究の参考」、「研究に再利用」、「研究を再現」は、全てデータ公開経験と関連がみられた（表 3-50）。

公開データを入手する際に問題があると認識していた回答者は 79.8%であった（表 3-15）。選択率が高かったのは、「利用料金」（43.1%）、「利用者登録が必要」（33.3%）、「利用条件がよ

3 章 日本の研究者によるデータ公開状況と内的要因

くわからない」(33.1%)の順であった(表 3-16)。Schmidt et al. (2016a)の調査で「明らかに負担である」の回答率が最も高かったのは²²⁾本研究と同様に「利用料金」(69%)であるが、2位以下は「データごとに品質が異なる」(53%)、「データの収録基準が異なる」(50%)、「データごとにフォーマットが異なる」(49%)であった。本研究では2位の「利用者登録が必要」は10位(18%)、3位の「利用条件がよくわからない」は9位(22%)であり、差がみられた。日本の研究者の回答は、どちらかという利用以前の問題点を選択する回答者が多く、Schmidt et al. (2016a)の調査結果では、公開データを繰り返し利用した経験による問題点を選択する回答者が多いという傾向がみられた。

公開データの検索に用いられているツールや情報源は、「サーチエンジン」(83.6%)、「論文や学術記事の参考文献」(71.0%)が多く、次いで「出版社や学術雑誌のサイト」(37.8%)の順であった。サーチエンジンや参考文献が上位を占めている点は先行研究と同様であった。しかし、Kratz and Strasser (2015a)の調査では2位が「研究コミュニティで使われているデータベース」であり、Schmidt et al. (2016a)の調査では3位が「データリポジトリでの検索」(63.2%)であったことから、これらの調査と比べてリポジトリやデータベースがあまり使われていないと考えられる。一方、論文は「サーチエンジン」(83.3%)、「論文情報のデータベース」(70.8%)、「参考文献」(61.2%)の順であり、データベース²³⁾がよく使われていた。

入手した公開データの判断基準として最も選択率が高かったのは、「著者情報」(71.5%)であった。一方、公開データの入手における問題のうち、「著作者情報がよくわからない」は20.3%が選択していた。データ公開の際には、著作者情報を明示またはORCIDなどの研究者データベースへのリンクを確立しておく、一層の利用に資すると考えられる。公開データの判断基準として次いで選択率が高かったのは、「研究手法の確かさ」(63.3%)、「そのデータを用いた論文」(58.7%)であった。論文については「掲載雑誌」(87.4%)が最も選ばれていた。データの「掲載リポジトリ」は、論文の「掲載雑誌」にあたると考えて選択肢としたが、選択率は25.8%にとどまった。雑誌と同様に信頼性の高いリポジトリが増えることによって、利活用が進むのではないかと考えられる。

今後、公開データを利用してみたい分野については、論文や、これまで利用したことのあるデータ以上に多くの分野が選択されていた(平均2.7分野)。このことは、研究者は他者のデータを利用したいとするTenopir et al. (2011)などの結果と一致していた。前述した公開データの入手における問題の解決や検索ツールの周知によってデータ入手が容易になれば、公開データの利用が促進されるのではないかと考えられる。

²²⁾ Schmidt et al. (2016a)は、各項目を「明らかに負担である」、「やや負担である」、「負担ではない」の3件法で尋ねており、選択肢は11項目である。また、結果は小数点以下の数値が示されていないかった。

²³⁾ 選択肢では、Web of Science, Scopus, CiNii Articles を例示した。

3.3.4 日本の研究者が扱うデータのプロフィール

3.1.3 で述べたように「内閣府報告書」などでは、研究者が扱うデータの種類に応じた配慮が必要であると指摘されている（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015）ことから、「研究に用いるデータの種類はデータ公開に影響している」という仮説を立てて、研究にデータを用いている回答者を対象として調査した。具体的には、(1)データ量が多い、(2)データの所有権が複数の人や組織にある、(3)データに機密情報が含まれる、(4)他の研究者によるデータの理解が難しいと考えている場合は、データを公開しない傾向にあるという仮説を立てて調査した。

しかし、データ公開率との関連を確認したところ、(1)から(3)、および(4)のうち、自身と同じ分野の研究者については差が認められず、全て棄却された。唯一、(4)異分野の研究者がデータを理解することは難しいと考えている回答者のみデータの公開率が高かった。「扱うデータによって、データ公開の困難さが異なるのではないか」と論じられることが多いが、本調査ではデータそのものよりも、研究者の懸念や属性の方がデータ公開に影響を与えている可能性が示唆された。なお、データの機密情報については「その他、守秘義務がある情報」を過半数（58.4%）の回答者が選択していたが（表 3-23）、機密に関わるため具体的な内容の入力を求めなかった。この点は今後の課題としたい。

3.3.5 分野による差異

分野によって、データ公開の状況やその理由、障壁、扱うデータなどに差がある場合とない場合があることがわかった。以下では、国内外の状況を参照しながら考察する。

[1] 分野別データ公開の状況

データ公開経験をもつ回答者の比率（以下、「データ公開率」）は分野によって差が大きく、公開率が高かったのは植物・動物学（84.2%）、分子生物学・遺伝学（70.5%）であった（図 3-6）。データ公開率が低かったのは精神医学・心理学（30.8%）、工学（41.1%）であった。“研究分野により取組に違いがある”（文部科学省科学技術・学術審議会学術分科会第8期学術情報委員会, 2016, p. 2）ことを実証する結果であったと考えられる。

データ公開の理由についても、分野による差がみられた（表 3-25）。科学的利他性に関する項目は、全体的に国外の調査と比較して選択率が低かったが、「科学研究や成果実装を推進したいから」や「オープンデータに貢献したいから」については、いずれも宇宙科学、植物・動物学、コンピュータサイエンスの選択率が高かった。一方、「他の研究者からのリクエストに応じて」については、全ての分野で選択されていた。

「分野・コミュニティの規範」は全体では7位（9.5%）であった。Kim and Stanton (2015)

や Berghmans et al. (2017) では主要な動機とされており、データ公開が推進されてきた分野では選択率が高いのではないかと推測される。しかし、選択率が上位の数学 (28.6%)、植物・動物学 (25.0%)、分子生物学・遺伝学 (22.6%) でも 3 割に満たなかった (表 3-25)。分野別に選択率を確認して他の理由との関連を調べたところ、8 位の「業績になる場合があるから」 (9.3%) と有意な相関がみられた (図 3-7)。つまり、規範として認識されている分野では、業績として評価されている可能性が示唆された。

データ公開に関する制度である雑誌、所属機関、助成機関のポリシーについては、雑誌のポリシーのみデータ公開率との関連がみられた。つまり、データ公開率が高い分野では、雑誌のポリシーを公開理由として選択する回答者の比率が高かった。所属機関のポリシーは、社会科学のみ選択率が高かった (80.0%)。また、助成機関のポリシーは、経済学・経営学などの 7 分野では、全く選択されていなかった。現状では雑誌、所属機関、助成機関の要求は、分野によって差があるといえる。今後、多くの分野でデータ公開の義務化が進む場合、データ公開状況に影響を与えと考えられる。

[2] 分野別データ公開の障壁

3.3.2 で述べたように、データの非公開理由は全体的に選択率が高い項目がなく、データ公開意思が「ある」、「ない」、「わからない」回答者によって選択率が高い、あるいは低い項目もなかった。分野別の結果においても、データ公開率と関連があるデータの非公開理由は特定できなかった (表 3-26)。つまり、データ公開率が高い分野や低い分野ほど、選択率が高い／低い項目は見当たらなかった。一方、後述するようにデータを公開する場合の資源の充足度や懸念については分野やデータ公開経験の有無による差がみられた。自由記述からは、公開するデータがない分野があること (3.2.8 [3]) などが明らかになった。この点は、3.3.10 (データ公開が適さない場合の検討) で改めて考察する。

データを公開する場合の資源の充足度は、データ公開用リポジトリと研究中のストレージに差がみられた。データ公開用リポジトリは、精神医学・心理学と地球科学の充足度が比較的低く、植物・動物学の充足度が比較的高かった (図 3-8)。また、データ公開率とも関連がみられ、データ公開率が高い分野ほど充足度が高かった。研究中のストレージは宇宙科学と精神医学・心理学の充足度が比較的低く、植物・動物学とコンピュータサイエンスの充足度が比較的高かったが (図 3-9)、データ公開率とは関連がみられなかった。また、その他の資源については分野による差がみられなかった。つまり、データの保存用ストレージ、時間、人材、資金は、分野によらない問題であることがわかった。

データを公開する場合の懸念は、誤解や誤用の可能性以外は分野による差がみられた。つまり、誤解や誤用の可能性については、分野によらず懸念とされていることがわかった。差がみられた項目のうち、機密・プライバシー情報への懸念が強い分野は商用利用される可能性への懸念も強かった (図 3-15)。免疫学と臨床医学はいずれの懸念も強く、植物・動物学

と宇宙科学はいずれの懸念も弱かった。なお、データ公開率と関連がみられた項目はなく、データ公開率が高い・低い分野に関連のある特定の懸念はなかった。

[3] 分野別データ公開と関連のある経験

データの提供経験と被提供経験は、いずれも分野による差がみられた（表 3-28）。また、分野別の提供経験と被提供経験には強い相関がみられた。しかし、データ公開率とは被提供経験に中程度の相関がみられたのみであったことから、データの提供・被提供を頻繁に行っている分野はデータ公開率が高いとまではいえないことがわかった。本調査ではその原因までは明らかにできなかったが、データ公開を行っている分野では、データを直接やり取りするよりも、公開することによって他の研究者の利用に供しているのではないかと考えられる。この点については今後の課題としたい。

公開データの入手経験も分野による差がみられた（表 3-24）。分野別の公開データの入手経験とデータ公開率には相関がみられ、分子生物学・遺伝学や植物・動物学は入手経験も公開率も高く、精神医学・心理学や臨床医学はいずれも低かった（図 3-17）。入手したデータの利用は、「再分析・再利用して自身の研究を行う」について分野による差がみられた（図 3-18）。「行う」と「たまに行う」の合計が最も多かったのは経済学・経営学であり、「行う」が最も多かったのは植物・動物学であった。精神医学・心理学の回答者は全く行っていないかった。データの入手経験とデータ公開率に関連がみられたことから、リポジトリの利用状況を確認した（表 3-29）。論文のサーバは全ての分野で利用されている一方で、データリポジトリは検索、入手、公開ともに全く利用している回答者がいない分野が存在した。データ公開率が高い分野では、データの入手や公開先として分野リポジトリを選択する回答者の比率が高かった。

[4] 分野別研究者が扱うデータのプロフィール

分野によってカレントデータの量に差はみられたものの（表 3-30）、分野別のデータ量とデータ公開率には相関がみられなかった。全体の結果（3.3.4）と同様に、データ量の多寡は、分野別にみてもデータ公開と関連がない可能性が示唆された。

また、機密情報の有無も分野による差がみられた（表 3-24）。分野別の機密情報がある回答者の比率は、データを公開する場合の「機密・プライバシー情報」への懸念の強さ、および「商用利用される可能性」への懸念の強さと相関がみられた。機密情報が「商用利用される可能性」への懸念を喚起しているのかどうかを確認することは、今後の課題としたい。オープンサイエンス関連の政策では、機密やプライバシー保護の問題を検討した上で、場合によっては制限事項を設けるとされている（国際的動向を踏まえたオープンサイエンス検討会, 2015; 内閣府, 2016; 文部科学省科学技術・学術審議会学術分科会第 8 期学術情報委員会,

2016)。本調査では、免疫学、精神医学・心理学、臨床医学の回答者が扱うデータには機密情報があり、かつ懸念が強いことが明らかになった。

3.3.6 データ公開と論文の OA の差異

[1] 公開率

データ公開と OA の経験を比較すると（表 3-7）、データ公開経験は「あり」が 51.0%、「なし」が 46.9%、「わからない」が 1.9%であったのに対して、OA の経験は「あり」が 70.9%、「なし」が 24.8%、「わからない」は 4.3%であった。また、両方の経験をもつ回答者は 40.6%、OA 経験はあるがデータ公開経験はない回答者は 29.5%、データ公開経験はあるが OA 経験はない回答者は 8.8%であった。これらは OA の方がデータ公開よりも実践されている（OECD, 2015）ことを数値で裏付ける結果であったといえるだろう。

データ公開よりも OA の方が「わからない」の選択率が高かった要因として、回答者が意図せずに OA になっている場合があるからではないかと考えられる。なぜなら、公開理由の自由記述（表 3-9）で「消極的・強制的」に分類した「OA 論文に選出された（2 名）」、「知らない間に OA になっていた」といった回答がみられたからである。データについては、論文と同時に提出するとは限らず、公開する場合にはデータの整形や選択などが必要であるため、意図せずに公開される可能性は低いと考えられる。

OA 経験は分野による差が大きかったが（表 3-24）、データ公開経験との関連はみられなかった。つまり、OA の経験がある回答者の比率が高い分野ほどデータ公開率が高いとはいえなかった。たとえば、精神医学・心理学はデータの公開率が最も低かった（30.8%）が、全ての回答者（100.0%）が OA 経験を有していた。オープンサイエンスに関する政策では、分野による取り組みの状況や特性に応じた対応が必要であると指摘されているが（文部科学省科学技術・学術審議会学術分科会第 8 期学術情報委員会, 2016; 内閣府, 2017）、OA とデータ公開の実践状況に違いがある点にも留意が必要であろう。

[2] 公開理由

データ公開と OA の理由を比較すると（表 3-8）、データの 1 位は「研究成果を広く認知してもらいたいから（以下、「認知」）」（58.5%）、2 位は「雑誌のポリシーだから」（43.8%）であり、OA の 1 位は「雑誌のポリシーだから」（81.7%）、2 位は「認知」（46.9%）の順であった。データ公開と OA のいずれも公開による認知の向上と雑誌のポリシーが主な理由になっていた。より公開率が高い OA の理由が雑誌のポリシーであったことから、データ公開を求める雑誌が増加することによって公開率が上がるのではないかと考えられる。

「所属機関のポリシー」は、データは 5 位（23.7%）、OA は 6 位（8.3%）であった。大学

などの研究機関では、2015年に京都大学がOA方針を公開したことを嚆矢として(天野, 2016), OA方針の策定が徐々に広がっている。2018年10月現在, OA方針を公開しているのは25機関であるという²⁴⁾。データについては, 国立研究開発法人の海洋研究開発機構(2007)と同じく国立研究開発法人の国立環境研究所(2017)の2機関がデータポリシーを策定している。2018年6月15日に閣議決定された「統合イノベーション戦略」では, 「研究データの管理・利活用についての方針・計画の策定等」という項目が立てられ(内閣府, 2018, p. 8), 2018年6月29日には「国立研究開発法人におけるデータポリシー策定のためのガイドライン」が公開された(国際動向を踏まえたオープンサイエンスの推進に関する検討会, 2018)。こうしたガイドラインによって, 今後は研究機関によるデータポリシーの策定も増加する可能性があると考えられる。

「助成機関のポリシー」は, データは9位(6.3%), OAは8位(2.0%)であり, いずれも選択率が低かった。日本学術振興会の科学研究費助成事業はOAを推奨しているが(日本学術振興会, n.d.), 公開理由としてあまり選択されていなかった。また, 2018年にSpringer Nature社が日本の研究者を対象として実施した調査では, 回答者の38%が自身の主な助成機関のDMPに関するポリシーについて知らなかったと報告されていた(Springer Nature, 2018)。助成機関のポリシーの周知や認知度の向上も課題であると考えられる。

なお, 「所属機関のポリシー」と「助成機関のポリシー」は, いずれもデータの方がOAよりも選択率が高かった。しかし調査時点において, 日本ではデータ公開ポリシーよりもOAポリシーを制定している機関の方が多かったと考えられる。たとえば, 国外の助成機関や, 国外の共同研究者の所属機関のポリシーを指していた可能性があるが, 本調査ではそこまで確認することができなかった。これらの回答が実態に即しているのかどうか, あるいは質問が適切であったかどうかの検証を今後の課題としたい。

[3] 非公開理由と公開意思

論文のOA経験がない回答者407名に非公開理由を尋ねたところ, 雑誌のポリシーがないこと(60.4%)と資金(39.6%)に集中していた(表3-10)。そして非公開理由が解決した場合, 78.1%の回答者がOAを実践しても良いと考えていた(表3-11)。一方, データについては突出した理由がみられず, 雑誌のポリシーではない(26.4%), 時間が必要(25.9%), 所属機関にポリシーがない(22.4%)などが比較的選ばれていた(表3-10)。非公開理由が解決した場合でも, データ公開を実践しても良いと考える回答者は28.4%にとどまった(表3-11)。このようにOAとデータ公開は, 非公開理由とそれが解決した場合の公開意思に大きな違い

²⁴⁾ オープンアクセスリポジトリ推進協会(JPCOAR)によれば, 2018年10月末現在, 25機関がOA方針を公開している。“オープンアクセス方針・実施要領 リンク集”. JPCOAR. https://jpcoar.repo.nii.ac.jp/index.php?page_id=53, (accessed 2018-10-31).

3 章 日本の研究者によるデータ公開状況と内的要因

がみられた。現状では、データ公開経験がない研究者は、データ公開に対して論文の OA よりも慎重であることが示唆された。

[4] 入手と利用

データと論文の探索方法は、いずれもサーチエンジン（Google や Google Scholar など）が 1 位であった（表 3-17）。選択率は、データが 83.6%，論文が 83.5%であり、ほとんど差がなかった。2 位以下では「データ／論文のデータベース（以下、「DB」）」と「データ／論文のサーバ、リポジトリ（以下、本項のみ「リポジトリ」と記す）」に差があり、DB はデータが 6 位（18.1%）で論文は 2 位（71.0%），リポジトリはデータが 12 位（2.6%）で論文は 5 位（48.4%）であった。つまり、論文は DB やリポジトリで探索することが多いのに対して、データはこれらがあまり活用されていない状況にあると考えられる。現状では DB やリポジトリ自体が少なく認知度が低い段階にあるといえる。

データの信頼性の判断基準も、データと論文で差がみられた（表 3-18）。データの 1 位は「著者情報」（71.5%）であり、論文の 1 位は「掲載雑誌」（87.4%）であった。「データを掲載しているリポジトリ」は、論文の掲載雑誌に相当すると仮定して選択肢に含めたが、4 位（25.8%）にとどまった。信頼性の高いリポジトリも一定数存在すると考えられるが、その数は多くはないと推測される。データの 2 位は「研究手法の確かさ」（63.3%）であり、論文と同程度であった（64.6%）。3 位は「そのデータを用いた論文」（58.7%）であった。3.3.3 でも述べたように、1 位から 3 位の情報（著者、手法、論文）を確認するために、著者情報や論文への永続的なリンクを確立しておくと考えられる。

3.3.7 属性による差異

[1] 所属機関

所属機関によってデータの公開率や障壁に差があることがわかった。データの公開率は、回答者 1,398 名のうち、大学（56.9%）や公的機関・団体（55.6%）の回答者が高かった。しかし、大学の研究者は人材（図 3-19）、時間（図 3-20）、資金（図 3-21）の充足度が低いと認識しており、「引用せずに利用される可能性」（図 3-24）や「先に論文を出版される可能性」（図 3-25）、すなわちキャリアリスクへの懸念が強かった。次章で述べるように、データ公開を業績として評価することによって、キャリアリスクへの懸念が低減すると考えられる。

企業の研究者は、基本的に公的資金によらず営利を目的として研究活動を行っているが、それでも 33.9%がデータ公開経験を有していた。ただし、商用利用される可能性や機密・プライバシー情報への懸念は強く、カレントデータの所有権をもつ人や組織（88.8%）、機密情報（70.5%）の比率が大学や公的機関・団体の回答者と比較して高かった。自由記述では、特許

を取得するため公開は難しいとする回答が多数みられた。企業の利益保護は当然のことであり、特許に関わるデータは公開する必要はないと考えられる。また、所有権が産学にわたるデータがあることも確認されたため、共同研究やデータ提供が行われていることが推察された。

[2] 年齢

年齢については、1,398名の回答を5年ごとに集計した上で、年齢層別に差があるかどうかを確認した。データ公開率は、年齢層による差はみられなかった。つまり、年齢層が高い方が、あるいは低い方が、公開率が高いといった傾向はみられなかった。

データを公開する場合の懸念のうち、「先に論文を出版される可能性」(図 3-27)と「引用せずに利用される可能性」(図 3-28)、すなわちキャリアリスクへの懸念は年齢層による差が大きかった。Tenopir et al. (2015)やSchmidt et al. (2016a)の結果と同様に、キャリア形成期である若年層の懸念が強かった。その他の懸念やデータを公開する場合の資源の充足度については、年齢層による差はみられなかった。

公開データの入手経験は年齢層による差があり、年齢層が高くなるほど提供や被提供の頻度が高くなる傾向がみられた。キャリアが長い研究者ほど、データや人脈の蓄積があり、データのやり取りを行うことが増えるのではないかと考えられる。

3.3.8 データ公開経験の有無による差異

3.2.10に示した結果から、データ公開経験の有無によって、障壁の程度などが異なることがわかった。データを公開する場合の資源のうち、充足度が高いほどデータ公開経験がある回答者の比率が高かったのは、人材(表 3-35)、時間(表 3-36)、資金(表 3-37)公開用のリポジトリ(表 3-39)、であった。また、データを公開する場合の懸念は、「機密・プライバシー情報」(表 3-44)と「商用利用される可能性」(表 3-45)に差がみられ、公開経験がある回答者のほうが弱かった。3.3.2では、データ公開経験がない回答者の非公開理由について考察し、公開意思と関連する非公開理由が特定できなかったと述べた。しかし、データ公開経験の有無による差から、リポジトリや時間などの資源の充足度や商用利用される可能性、機密・プライバシー情報への懸念が強いことがわかった。これらの要因がデータ公開意思に影響を与えているのかどうかの検証を今後の課題としたい。

データ公開経験がある回答者は、データを提供する頻度(表 3-47)、提供される頻度(表 3-48)、公開データを入手する頻度(表 3-49)、および入手したデータを「研究の参考」、「再分析・再利用」、「再現・追試」する頻度(表 3-50)が高く、公開データの利用を希望する比率も高かった(表 3-51)。一方、データのプロファイル、すなわちデータの量や機密情報の有無については、データ公開経験の有無による差はみられなかった。先行研究では、データの

提供経験や利用経験がデータ公開に繋がると論じられていた (Damvad, 2014; Berghmans et al., 2017)。本調査ではデータ公開経験とこれらの経験に関連がみられたが、その因果関係までは明らかにできなかった。この点も今後の課題としたい。

3.3.9 データ公開に消極的な研究者の特徴

3.2.11 では、データ公開経験がなく、かつ、何らかの非公開理由をもつ回答者 (595 名) のうち、データの非公開理由が解決した場合にデータ公開の意思がある回答者 (169 名) とない回答者 (130 名) の差異を確認した²⁵⁾。ここでは、データの非公開理由が解決した場合であっても、データ公開を行わないとする回答者 (以下、「データ公開に消極的な研究者」) に着目して、その特徴について考察する。

データ公開に消極的な研究者は、データを公開する場合の資源については公開意思がある回答者との差がみられなかった。データを公開する場合の懸念は、「先に論文を出版される可能性」(図 3-32)、「誤解や誤用の可能性」(図 3-33)、「商用利用される可能性」(図 3-34) が強かった。データの提供経験 (図 3-35)、被提供経験 (図 3-36)、公開データの入手経験 (表 3-59)、入手したデータの再分析・再利用経験 (図 3-37) は、いずれもデータ公開意思がある研究者の方が消極的な研究者よりも頻度が高く、公開データの利用を希望する回答者の比率も高かった (表 3-60)。データのプロフィールについては、差はみられなかった。

以上のことから、データ公開に消極的な研究者は、そうでない研究者と比較して、先に論文を出版される可能性などへの懸念が強く、データの共有や入手、利用も比較的行っていないことがわかった。ただし、データの量や機密情報の有無などは、関連がみられなかった。

3.3.10 データ公開が適さない場合の検討

ここではデータ公開が適さない場合について検討する。まず、0 (データの非公開理由と公開意思) に示した自由記述のうち、(a)不要、(c)公開するデータがないという指摘からデータ公開の必要性について考察する²⁶⁾。続いて、分野や所属機関による事情についても検討する。

(a)データ公開は不要であるとする意見のうち、「論文や補足資料で充分である」について、これを支持する調査結果として補足資料の有用性が示されている (Anagnostou et al., 2015; Kafkas, Kim, Pi, & Mcentyre, 2015)。その一方で、補足資料は再利用には適さないこと (Kenyon, Sprague, & Flathers, 2016) やリンク切れ (van den Eynden & Bishop, 2014; Evangelou et al., 2005; Anderson et al., 2006; Pepe et al., 2014)、質の問題 (Jubb, 2013) などが指摘されていることか

²⁵⁾ 3.2.11 および本項では、「わからない」を選択した 296 名は除外した。

²⁶⁾ (b)公開のインセンティブの欠如については[1]で述べる。

ら、補足資料のみで充分とは言い難いと考えられる。また、「必要ならば著者に直接問い合わせれば良い」という意見もみられたが、植物学と動物学を対象とした調査によって出版後 20 年が経過した論文の約 8 割は著者に問い合わせてもデータが入手できないことが明らかにされている (Vines et al., 2014)。以上のことから、基本的にはデータを公開して長期にわたる再利用に供することとして、論文や補足資料で十分な場合については別途検討することが望ましいと考えられる。「必要性が感じられない」とする意見について、研究者の立場からは公開の意義や再利用の可能性、保存の必要性が感じられないといった見解が示されていた。一方、オープンサイエンス政策 (OECD, 2015; 国際的動向を踏まえたオープンサイエンスに関する検討会, 2015; 日本学術会議, 2016) では、現時点では想定されていないイノベーションの創出や分野を超えたデータの利活用などが期待されている。実際に、表 3-19 (データを利用してみたい分野) の結果からも、研究者は異分野のデータを利用してみたいと考えていることが示唆された。このギャップを埋めるためには、データの再利用によるベストプラクティスや経済的な効果などの具体的な将来像を示す必要があると考えられる。

(c)公開するデータがないことについては、理論研究を行う純粋数学など分野による事情についての自由記述がみられた。また、“(物性物理学は) 生データを公開しても他研究者に理解することは難しい。理解できるように整えるのは、時間的にかなり負担である”との指摘もみられた。実際に、「異分野の研究者によるデータの理解の難しさ」(図 3-16) では、物理学の回答者は「難しいと思う」(43.1%) と「やや難しいと思う」(49.0%) の選択率が高く、全体では免疫学、精神医学・心理学に次ぐ 3 位であった。こうした分野については別途調査と検討を行った上で、一律の公開を求めないといった対応が必要であると考えられる。

このほか、企業の研究者は機密・プライバシー情報と商用利用される可能性への懸念が特に強く、特許のため公開は困難であるという意見が多数みられた。分野別では、臨床医学や免疫学の研究者はこれらの懸念が強かった。また、“工学は製品やノウハウなど企業秘密に近いため”公開が困難であるとの指摘がみられた。第 5 期科学技術基本計画では、国家安全保障等に係るデータ、商業目的で収集されたデータを公開適用対象外とし、個人のプライバシーや財産的価値があるデータについては制限事項を設けるとしている (内閣府, 2016, p. 32)。図 1-3 (データの公開レベル) の「クローズ」に相当する情報を含むデータについては、公開する必要がないというコンセンサスを得るため、具体例などをわかりやすく示す必要があると考えられる。

3.4 まとめ

本章では、日本の研究者によるデータ公開の状況と内的要因 (動機と障壁) を明らかにすることを目的として、質問紙調査を実施した。調査対象は、NISTEP 科学技術動向研究センターが運営する科学技術専門家ネットワーク (1,983 名) であり、大学、企業、公的機関・団体

3 章 日本の研究者によるデータ公開状況と内的要因

の研究者から回答を得た。調査期間は 2016 年 11 月 30 日から 12 月 14 日である。有効回答 1,406 名 (70.9%) のうち、現在、研究活動(口頭発表や論文出版)を行っている 1,398 名 (70.5%) の回答を分析対象とした。

データ公開経験をもつ研究者(以下、「データ公開率」)は 51.0%であり、大学の研究者に限ると 56.9%であった。2014 年の Wiley 社の調査ではデータ公開率が 52%であった(Ferguson, 2014)のに対して日本の研究者は 44%であった。また、2016 年の CWTS/Elsevier の調査では、データ公開率が 66%であった (Berghmans et al., 2017)。つまり、日本の研究者のデータ公開率は国外の研究者と比較するとやや低い可能性が示唆された。論文の OA 率は 70.9%であり、データ公開よりも OA の方が実践されていることを実証する結果であった。データ公開の実践方法は、個人や研究室のウェブサイト、論文の補足資料が多く、パブリックリポジトリでの公開は限定的であった。これは、国外の調査と同様の傾向であるといえる。

データ公開の主な理由(動機)は、研究成果の認知向上や雑誌のポリシーなどであった。国外の調査や小野ら(2016)の調査で主要な公開理由とされていた科学的利他性(科学研究や成果実装の推進、他の研究者のリクエストに応じて、オープンデータへの貢献)は、選択率が低かった。制度のうち、所属機関や助成機関のポリシーは選択率が低かった。助成機関の一部はデータマネジメントプランの提出を求めるようになったため(科学技術振興機構 [JST], 2017)、今後はその影響がみられると予想される。

データの非公開理由は突出した項目がなく、雑誌のポリシーではない(26.4%)、時間が必要(25.9%)、所属機関にデータ公開方針がない(22.4%)、ニーズがないと思う(22.1%)など多岐にわたった。自由記述では、データ公開のインセンティブがないことなどが指摘された。また、これらの理由が解決した場合、データを公開するという回答者は 28.4%に留まり、わからないとする回答者は 49.7%であった。一方、論文の非公開理由は雑誌のポリシーではない(60.4%)と資金が必要(39.6%)に集中しており、非公開理由が解決した場合は公開するという意思をもつ回答者の比率がデータ(28.4%)と比べて高かった(78.1%)。

データ公開の障壁として、データを公開する場合の資源(人材、資金、時間、データ保存用ストレージ、データ公開用リポジトリ、研究中のデータストレージ)の充足度について尋ねたところ、全体的に不足感が強かった。また、データ公開に対する懸念(引用せずに利用される可能性や商用利用される可能性など)について尋ねたところ、「研究の誤りを発見される可能性」を除き、懸念も強かった。

データ公開に関連があると考えられる項目(データ提供経験や公開データの利用経験など)についても調査を行い、データ公開経験の有無やデータ公開意思の有無との相関を調べた。結果を表 3-61 に示す。質問項目ごとに、分野、属性(所属、年齢)、データ公開経験、データ公開意思によって差がみられた場合は「○」を付した。分野は 16 項目、所属は 17 項目、年齢は 9 項目、データ公開経験は 15 項目、データ公開意思は 9 項目について、差がみられた。

表 3-61 質問項目と分野・属性・データ公開経験・データ公開意思との関連

質問項目	分野	所属	年齢	公開	意思
Q18 データの公開経験	○	○		—	—
Q2 論文の公開（OA）経験	○	○	○	○	
カレントデータを公開する場合の障壁の程度					
Q28 公開の資源	○				
研究中のストレージ	○				
公開用リポジトリ	○			○	
保存用ストレージ					
時間		○		○	
人材		○		○	
資金		○		○	
Q29 データ公開の懸念	○	○		○	○
商用利用	○	○		○	○
機密・プライバシー	○	○		○	
誤解や誤用					○
引用せずに利用	○	○	○		○
研究の誤りを発見	○	○			
先に論文を出版	○	○	○		○
データの提供・被提供状況					
Q9 データの提供頻度	○	○	○	○	○
Q10 データの被提供頻度	○		○	○	○
公開データの入手経験					
Q11 公開データの入手経験	○		○	○	○
Q12 公開データ入手における問題		○			
Q13 公開データの利用目的				○	
研究の参考				○	
再分析・再利用	○	○		○	○
再現・追試		○		○	
Q17 公開データの利用希望				○	○
カレントデータのプロフィール					
Q22 データ保存期間規定の有無		○	○		
Q23 データの量	○		○		
Q24 データの所有権		○			
Q25 データの機密情報	○	○	○		
Q30 データの理解（自分分野）					
Q31 データの理解（異分野）	○			○	

「公開」=データ公開経験, 「意思」=データ公開意思

3 章 日本の研究者によるデータ公開状況と内的要因

以上のように、本章では日本の研究者によるデータ公開状況と内的要因を調査して、論文の OA や国外の調査結果と比較した。また、分野、所属機関、年齢による違いを分析した。次章では、データ公開の要求状況を明らかにするために、学術雑誌によるデータ公開ポリシーを調査する。その上で、日本の研究者がデータ公開要求にどれだけ応えられているのかを明らかにするために、分野別にデータ公開ポリシーの掲載率とデータ公開率を比較する。

4 章 データ公開の外的要因と日本の研究者の状況

日本の研究者にとってのデータ公開の主要な外的要因として、学術雑誌がどの程度データ公開を要求しているのかを明らかにするために、学術雑誌（以下、「雑誌」）のデータ公開ポリシー（以下、「ポリシー」）を調査した（池内と逸村, 2016）。具体的には、2014 年 4 月から 6 月にかけて、22 分野各 10 誌、合計 220 誌の雑誌の投稿規定などを調査して、ポリシーの掲載率と要求の強度（公開を必須とするものから公開してもよいとするものまで）、および掲載誌の特徴を明らかにした。結果のデータセットは figshare¹⁾で公開している（Ikeuchi, 2016）。

続いて、データ公開要求に対して日本の研究者はどの程度データを公開しているのかを明らかにするために、本章で調査したポリシーと 3 章で明らかにした日本の研究者によるデータ公開率を分野別に比較した。本章は、4.1 で調査・分析方法を、4.2 で結果を、4.3 で考察を、4.3.5 でまとめを述べる。

4.1 方法

先行研究によるポリシーの強度の分類方法や調査対象誌の特徴は多様であった。そこで、最も詳細な方法に準じて予備調査を行い、適切な方法や対象を検討した。その上で、分野別にポリシーの状況を調査することによって、分野によるデータ公開の要求の度合いを明らかにした。また、どのような雑誌がポリシーを掲載しているのかを明らかにするために、先行研究（2.5）と予備調査の結果を参考に、ポリシーと関連があると考えられる特徴について仮説を立てて調査した。以下では、予備調査の概要と本調査の方法について述べる。

4.1.1 予備調査

ポリシーの調査・分析方法を検討するために、2.5 に示した先行研究を参考に予備調査を実施した（池内, 2013）。調査対象は 18 分野各 10 誌、合計 180 誌とした。調査期間は 2013 年 10 月 1 日から 30 日である。本項では、予備調査の方法と結果、および結果に基づく調査方法の修正点を示す。

¹⁾ figshare（リポジトリ）. <https://figshare.com>, (accessed 2018-10-31).

[1] 調査方法

ポリシーの区別について、先行研究では、リポジトリなどに登録する方法と補足資料で公開する方法を区別している調査と、両者を区別しない調査がみられた（表 2-9）。予備調査では、両者を区別することとして、リポジトリにデータを公開して論文に識別子を記すポリシー（Repository Policy, RP）と、補足資料にデータを掲載するポリシー（Supplemental Materials' Policy, SP）の2種類について調査を行った。なお、論文に関する補足的なデータを外部のリポジトリに公開して、論文の補足資料のセクションには識別子のみを記載する場合は SP とみなした。

[2] 調査対象誌の選定

調査対象誌は各分野のインパクトファクター（Impact Factor, IF）が高い雑誌とした。その理由を以下に述べる。

オープンサイエンス政策では、国際動向に対応することを志向している。たとえば、「科学技術イノベーション総合戦略 2017」では“国際的なルールメイキングに対応していくことも重要”（内閣府, 2017, p. 86）であると述べられている。文部科学省科学技術・学術審議会学術分科会学術情報委員会（2016）は、“研究データの範囲とその様式については、国際的な動向や原則公開とする趣旨を踏まえた上で、学協会等において、研究上の必要性等を考慮した検討”（p. 8）を行うとし、“国際的な動向も踏まえ着実に取組を進めることが求められる”（p. 2）と指摘している。ここでいう「国際的な動向」を把握するためには、現在、日本の研究者がよく利用している雑誌に限らず、国際的に注目度が高い雑誌を対象とすることが適切であると考えた。

また、1.3.1 で述べたように、第5期科学技術基本計画は、被引用数が Top10%の論文を増加させることを目標としている（内閣府, 2016, p. 30）。現在、日本の Top10%論文の雑誌カテゴリ（国内・国外のオープンアクセス（Open Access, OA）誌・非 OA 誌）ごとの全論文数に占める比率は、国内誌（OA 誌 3.7%, 非 OA 誌 1.5%）よりも国外誌（OA 誌 11.6%, 非 OA 誌 13.4%）の方が高いという結果が示されている（福澤, 2016）ことから、注目度の高い国外誌の動向を把握し、日本の研究者の公開状況と比較するのは意味があると考えられる。

データ公開の目的として、研究のグローバル化（国際共同研究の推進）や異分野データの活用が掲げられている。国や地域、分野を超えた研究成果を発表する場として、国際的に注目度が高い雑誌を調査して、日本の実態と比較することとした。IF の高さが分野のトップジャーナルと一致しない分野、たとえば引用されるまでの期間が長い分野、あるいは会議録や単行書が重視される分野があることも指摘されているが（大久保と赤井, 2005）、多分野の雑誌を一貫した基準で選ぶためには IF を使用するのが適当であると判断した。

本調査の目的は、雑誌のポリシーを調べて分野横断で比較し、データ公開の要求状況を明

らかにすることである。しかし、表 2-9 に示したように、分野によってポリシーの掲載率が異なることが示唆されていたことから、これまで調査が行われていない分野、特にデータ公開が積極的に行われていない分野ではポリシーがほとんど、あるいは全く掲載されていない可能性がある。一方、生物学 (Piwowar & Chapman, 2008)、計算科学 (Stodden et al., 2013)、社会学 (Zenk-Möltgen & Lepthien, 2014) 分野においては IF が高い雑誌ほどポリシーの掲載率や強度が高いという研究結果が示されていた。

以上の理由から Thomson Reuters 社²⁾の Journal Citation Reports (JCR) の 2012 年版を用いて、分野ごとに IF の高い学術雑誌を上位 10 誌ずつ抽出して調査対象とすることとした。

JCR に収録された雑誌には、Web of Science Core Collection (WoS) の主題カテゴリ (Subject Category) が付与されており、分野間比較や分野の規格化にしばしば用いられている (Leydesdorff & Bornmann, 2015)。しかし、Science Citation Index (SCI) は 176 カテゴリ、Social Sciences Citation Index (SSCI) は 56 カテゴリと多数のカテゴリがあり、上位カテゴリと下位カテゴリのような階層はない。そこで、科学工学指標 (Science and Engineering Indicators, S&EI) の分類を参考にしながら³⁾カテゴリを 18 分野に集約した。以下では、図 4-1 に示した自然科学分野と社会科学分野の分類、および調査対象誌の抽出手続きについて述べる。

²⁾ 2016 年 10 月から Clarivate Analytics 社。

³⁾ このほか、英国高等教育統計機構 (Higher Education Statistics Agency, HESA)、米国国立教育統計センター (National Center for Education Statistics, NCES)、日本学術振興会科学研究費助成事業、総務省統計局科学技術研究調査報告の分野分類などを検討したが、S&EI の分類の粒度と範囲が最も適切であると判断した。

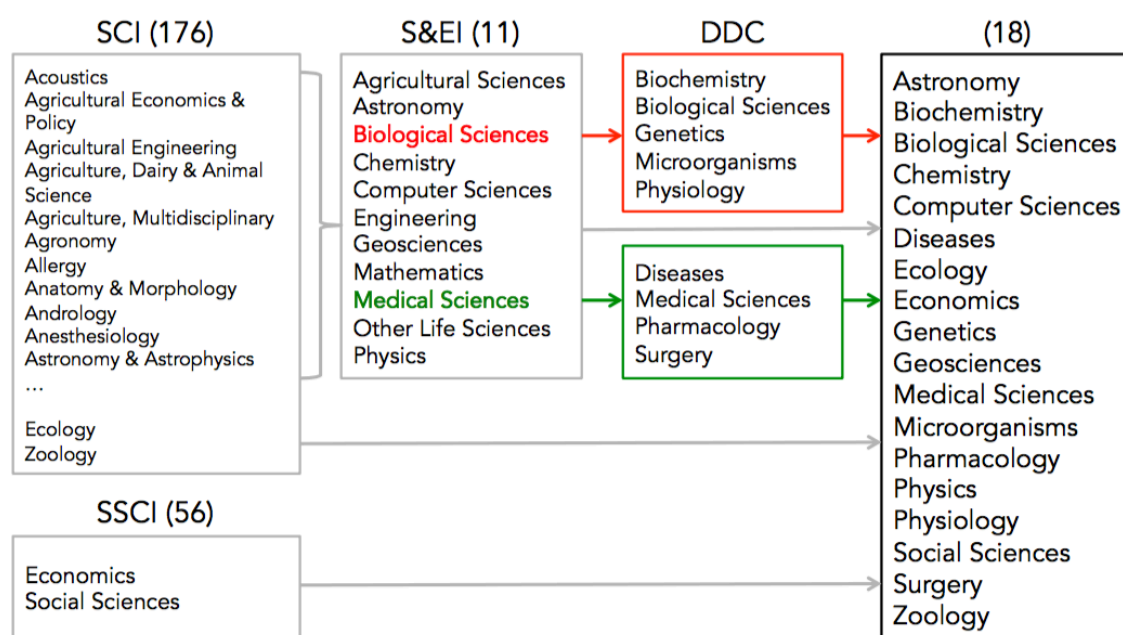


図 4-1 18 分野の分類と雑誌の選択（予備調査）

自然科学分野は、まず、SCI の 176 カテゴリーを、S&EI の 13 分野から社会科学と心理学を除いた 11 分野に分類した。さらに、データ公開が進んでいるとされている生物学 (Biological Sciences) と医学 (Medical Sciences) については下位区分を設けて細分し、分野内での比較を可能にすることとした。しかし、S&EI の下位区分は生物学が 12 区分、医学が 34 区分であり、詳細になりすぎる。そこで、生物学に相当する JCR の 22 カテゴリーと医学に相当する 43 カテゴリーをそれぞれ統合して、2 分野の IF 上位 500 誌を抽出した。合計 1,000 誌分のデューイ十進分類番号 (Dewey Deciaml Classification Number, DDC) を Ulrichsweb.com から取得し、これを基に下位区分として分類した。以上の手続きによって得られた 30 分野から、分野あたりのタイトル数や分野の散らばりを勘案して、天文学、生化学、生物学、化学、コンピュータサイエンス、疾患学、生態学、遺伝学、地球科学、医学、微生物学、薬理学、物理学、生理学、外科学、動物学の 16 分野を調査対象として選定した。社会科学分野は、SSCI の 56 カテゴリーから、公開された社会調査データや統計データの 2 次分析が行われている社会学 (佐藤, 石田, 池田, 2000; Borgman, 2015) と、先行研究で調査が行われていた経済学を対象とした。

以上の手続きによって、自然科学 16 分野と社会科学 2 分野、計 18 分野から IF の高い雑誌を各 10 誌ずつ選択した。なお、レビュー誌や速報誌などは、その性質上、データ公開を求めるとは考え難いため除外して、原著論文を中心に掲載している雑誌 (以下、「原著論文誌」) のみを選択することとした。また、単行書 (モノグラフ) シリーズも除外した。雑誌の種類は、タイトルではなく、雑誌の概要 (Overview など)、編集方針 (Aims & Scope など)、投稿

規定、掲載論文を確認することによって判断した⁴⁾。

合計 180 誌のウェブサイトで投稿規定や執筆要綱など（以下、まとめて「投稿規定」と記す）を通読して、データ公開に関する記述、すなわちポリシーを調査した。投稿規定に記述が見当たらない場合は、“data”をキーワードとしてサイト内検索を行った。より厳密な調査を行うためには、Gherghina and Katsanidou (2013) のようにウェブサイトを調査した後に、電子メールで雑誌の編集者に尋ねるという方法も考えられる。しかし当該調査によれば、電子メールの回答率は 44.1%（102 誌のうち 45 誌）と低く、対象者の負担になると考えられ、電子メール調査によってポリシーを有していることが新たにわかったのは 1 誌のみであった。以上のことから、本調査では電子メールによる質問は行わないこととした。

ポリシーによる要求の強度は、最も詳細な調査を行っていた Stodden et al. (2013) を参考として、5 段階に分類して調査した。表 4-1 に、ポリシーの強度と、判断基準とした表現の例を示す。「必須（査読・出版条件）」は査読や出版の条件としてデータ公開を要求する記述、「必須」は査読や出版の条件とは書かずに要求する記述、「推奨」はデータ公開を勧める記述がみられた場合とし、「言及のみ」はデータ公開について記載しているが特に公開を要求したり勧めたりしていない場合、そして「記載なし」はデータ（公開）に関する記述がない場合とした。ポリシーの対象となるデータの種類によって強度が異なる場合は、より高い方を採用した⁵⁾。

表 4-1 データ公開ポリシーの強度の分類（予備調査）

要求の強度	表現例
必須（査読・出版条件）	require, must, should, condition of publication
必須	require, must, should
推奨	encourage, recommend
言及のみ	(implied)
記載なし	(no mention)

[3] 結果

全ての調査対象誌は電子版、または電子版と印刷版の両方が存在しており、ポリシーをウ

⁴⁾ たとえば、原著論文が中心である *Ecology Letters* や、同じく原著論文が中心で定期刊行される *Ecological Monographs* は対象とし、レビュー論文やエッセイが中心の *Angewandte Chemie* や、化学者が最新のトピックを提供することを目的とした *Aldrichimica Acta* は除外した。

⁵⁾ たとえば、*Journal of the American Chemical Society* は、物質によって要求の強度が異なり、原子座標はタンパク質データバンクへ登録するべき (should be deposited) とされているが、粉末回析データは推奨 (encouraged) とされている。こうした場合は、「必須」にカウントした。

4 章 データ公開の外的要因と日本の研究者の状況

ウェブサイトで確認することができた。表 4-2 に、予備調査の集計結果を示す。排列は、RP の強度が高いタイトル数が多い順とした。

表 4-2 分野別学術雑誌のデータ公開ポリシー：予備調査 (n=180)

分野	リポジトリ (RP)						補足資料 (SP)					
	あり	*必須	必須	推奨	言及	なし	あり	*必須	必須	推奨	言及	なし
微生物学	10	7	3	0	0	0	9	2	0	1	6	1
生理学	9	7	2	0	0	1	10	5	1	3	1	0
生物学	9	6	3	0	0	1	8	2	1	4	1	2
生化学	10	6	2	2	0	0	10	6	0	1	3	0
遺伝学	9	6	2	1	0	1	9	4	0	1	4	1
物理学	6	5	1	0	0	4	8	3	0	1	4	2
医学	10	4	5	1	0	0	10	2	0	0	8	0
化学	7	4	3	0	0	3	7	2	2	2	1	3
地球科学	8	3	3	2	0	2	10	4	1	3	2	0
生態学	8	3	2	2	1	2	10	1	0	0	9	0
疾患学	7	1	6	0	0	3	9	1	0	2	6	1
動物学	8	1	3	4	0	2	9	0	0	1	8	1
薬理学	3	1	2	0	0	7	10	1	0	1	8	0
天文学	3	1	0	2	0	7	7	1	0	4	2	3
経済学	3	0	3	0	0	7	9	6	3	0	0	1
CS	1	0	1	0	0	9	3	0	0	1	2	7
外科学	0	0	0	0	0	10	4	0	0	1	3	6
社会学	0	0	0	0	0	10	2	0	0	0	2	8
合計	111	55	41	14	1	69	144	40	8	26	70	36
(%)	61.7	30.6	22.8	7.8	0.6	38.3	80.0	22.2	4.4	14.4	38.9	20.0

*査読・出版の条件とする

RP は、180 誌中 111 誌 (61.7%)、SP は 144 誌 (80.0%) に掲載されており (強度にかかわらずポリシーが掲載されている比率、表 4-2 の「あり」)、SP の方が掲載率が高かった。RP で最も多く掲載されていたのは「必須 (査読・出版条件)」(30.6%) であり、SP は「言及」(38.9%) であった。

ポリシーの掲載率および要求の強度は分野によってばらつきがみられた。RP の掲載率と要

求の強度が最も高かったのは微生物学（「必須（査読・出版条件）」7誌、「必須」3誌）であり、外科学と社会科学は全く掲載されていなかった。SPは全ての分野で複数誌に掲載されていたが、同じく分野による差がみられた。SPの掲載率と要求の強度が最も高かったのは生化学（「必須（査読・出版条件）」6誌、「推奨」1誌、「言及」3誌）であり、最も低かったのは社会学（「言及」が2誌のみ）であった。

予備調査から得られた主な知見は、次の8点である。

- (1) RPの掲載率は61.7%、SPの掲載率は80.0%であった。RPは「必須（査読・出版条件）」の比率が30.6%、SPは「言及」の比率が38.9%で最も高かった。
- (2) 生物学分野（微生物学、生理学、生物学、生化学、遺伝学）は、ポリシーの掲載率や要求の強度が高い傾向にあった。
- (3) 医学分野（医学、疾患学、薬理学、外科学）は、ポリシーの掲載率や要求の強度に差がみられた。たとえばRPの掲載誌数は、医学は10誌全て、疾患学は7誌、薬理学は3誌、外科学は0誌であった。
- (4) データ公開ポリシーの強度とIFの無相関検定の結果は、有意であった（ $r=0.529$, $p<0.05$ ）。外科学と社会学は、IFが高い雑誌を選択したにもかかわらず、10誌ともRPが掲載されていなかった。
- (5) OA誌（13誌）は、全てRPを「必須（査読・出版条件）」または「必須」としていた。
- (6) 同一の雑誌であっても、RPとSPの掲載の有無や要求の強度には差がみられた。
- (7) 研究の透明性を高めるために、研究倫理や利益相反（Conflict of Interest, COI）の開示について記載する雑誌がみられた⁶⁾。
- (8) データの公開先として、GenBankなどの具体的なリポジトリ名を指定、または推奨している雑誌や、“パブリックリポジトリ”などと指示している雑誌がみられた。

調査方法の問題点は、次の3点であった。

- (1) ポリシーの強度のうち、「必須（査読・出版条件）」と「必須」の違いは、Webサイトに掲載されている情報だけでは判別できない場合があった。すなわち、「査読・出版条件」という表現が含まれていない場合であっても、実際にはデータを公開しなければ査読・出版されない可能性が否定できない。従って、表4-2で「必須」にカウントした雑誌の中には、「必須（査読・出版条件）」にカウントすべき雑誌が含まれていた可能性がある。
- (2) 調査対象誌の選択手順が煩雑である。また、自然科学分野の調査対象誌について、SCI

⁶⁾ たとえば *Journal of Economic Literature* は、2012年1月1日以降、過去3年で合計で1万ドル以上の経済的支援を受けた場合は、その旨を明らかにするよう求めている。また、何らかの団体、委員、組織の長を務めている場合も開示する必要があるとしていた。

の 176 カテゴリを S&EI の 11 分野に分類する際に、調査者の主観が含まれる。

- (3) WoS の主題カテゴリは 1 雑誌あたり複数付与されているため、同一の雑誌が異なる分野で選択される場合がある。

以上の結果から、次の方針に従って本調査を行うこととした。詳細は 4.1.2 以降で述べる。

- (1) 医学分野が細分されている分類を採用する。また、1 雑誌あたり 1 分類が付与されているツールを用いる。
- (2) IF が低～中程度の雑誌を対象とすると、ポリシーが掲載されていない分野が増えると考えられるため、IF が高い雑誌を対象とする。
- (3) ポリシーを、RP と SP に分けて、それぞれ調査する。
- (4) 雑誌の特徴として、研究倫理や COI の開示に関する記載の有無を調査する。
- (5) ポリシーの強度は、「必須（査読・出版条件）」と「必須」を統合して 4 段階とする。

4.1.2 本調査の概要

予備調査の結果を受けて、まず、Thomson Reuters 社⁷⁾の Essential Science Indicators (ESI) の雑誌リストと JCR を用いて調査対象誌（22 分野各 10 誌、合計 220 誌）を選定した。続いて投稿規定などに書かれたデータ公開ポリシーの強度を雑誌のウェブサイトで調査した。あわせて雑誌の特徴を雑誌のウェブサイトや Serials Solutions 社の Ulrichweb.com で調査した。

本調査の期間は、ウェブサイトは 2014 年 4 月 8 日から 5 月 8 日、Ulrichweb.com は 2014 年 6 月 5 日である。予備調査と同様に、全ての調査対象誌は電子版のみか電子版と印刷版の両方が存在しており、ポリシーをウェブサイトで確認することができた。

4.1.3 調査対象誌の選定

4.1.1 予備調査の[2]調査対象誌の選定で詳述したように、(1)日本のオープンサイエンス政策は国際動向に対応することを目指していること、(2)日本の被引用数 Top10%論文の多くは国外誌に掲載されていること、(3)22 分野の雑誌を一貫した基準で選ぶ必要があることなどから、本調査においても国際的に注目度の高い雑誌として IF が高い雑誌を対象とすることとした。

調査対象誌の選定に用いた 2013 年 9 月版の ESI 雑誌リストには、11,872 誌が収録されており、各雑誌に自然科学と人文・社会科学の 22 分野が一意に付与されている（Thomson Reuters, 2013）。まず、ISSN がなく 2012 年版の JCR にも収録されていなかった 27 タイトルを除いて 11,845 誌を抽出した。続いて、IF を JCR（2012）から取得して、分野ごとに順位

⁷⁾ 2016 年 10 月から Clarivate Analytics 社。

付けを行った。

調査対象は、予備調査と同様に原著論文誌に限定して、レビュー誌、速報誌、短報誌、コメンタリー誌⁸⁾は除外した⁹⁾。雑誌の種類の判断も、予備調査と同様に雑誌の概要や実際に掲載されている論文に基づくこととした¹⁰⁾。

以上の手続きによって、22 分野の IF 上位 10 誌、合計 220 誌を抽出して調査対象とした。220 誌の異なり出版者数は 57 である。表 4-3 に、出版者ごとのタイトル数を示す。

表 4-3 出版者別タイトル数 (n=220)

出版社		出版社	
Wiley	31	American Physical Society	3
Elsevier	26	BioMed Central Ltd.	3
Nature Publishing Group (NPG)	25	Lippincott Williams & Wilkins	3
Springer	16	American Medical Association	2
Cell Press	11	American Meteorological Society	2
Oxford University Press	11	American Society for Microbiology	2
Institute of Electrical and Electronics Engineers	9	American Society of Plant Biologists	2
Institute of Physics Publishing	7	B M J Group	2
American Chemical Society	6	Cambridge University Press	2
R S C Publications	6	Cold Spring Harbor Laboratory Press	2
Taylor & Francis	5	Copernicus GmbH	2
Sage Publications, Inc.	4	Public Library of Science (PLOS)	2
The Lancet Publishing Group	4	Academy of Management ほか 31 社	1

出版者の表記は Ulrichsweb.com に準じた。

表 4-4 に出版国別のタイトル数を示す。調査対象誌の出版国は 13 か国であり、最も多かったのは米国 (104 誌)、次いで英国 (71 誌) であった。出版国は Ulrichsweb.com の「出版国」で調べたが、学協会などで国外の出版社から刊行している 7 誌については、学協会の国名や所在地を出版国とした。同じく、国際機関による 9 誌は本部や事務局の所在地を出版国とし

⁸⁾ *Behavioral and Brain Sciences* は、Neuroscience 分野で IF が 3 位であったが、一本のターゲット論文に対して複数の Open Peer Commentary を掲載するという趣旨の雑誌であったため、除外した。

⁹⁾ *Annual Reviews* のレビュー論文には補足資料を掲載することができ、わずかながら補足資料付きの論文も存在する。また、*Ecology Letters* には原著論文は掲載されていないが、データ登録ポリシーがある。しかし、レビュー誌全体からみればごくわずかであったため、これらは例外とみなして本調査の対象は原著論文誌に限定した。

¹⁰⁾ たとえば、原著論文を中心とした *Academy of Management Review* は調査対象としたが、*Frontiers in Neuroendocrinology* のように、タイトルに“review”が含まれていなくてもレビュー論文を中心とした雑誌は除外した。

た¹¹⁾。出版国別のリポジトリ数を付表 51 に示す。

表 4-4 出版国別タイトル数 (n=220)

出版国		出版国	
米国	104	フランス	2
英国	71	インド	1
ドイツ	13	日本	1
オランダ	13	ニュージーランド	1
中国	6	シンガポール	1
カナダ	3	スウェーデン	1
スイス	3		

本研究では、分野ごとに IF が高い雑誌を抽出して調査を行ったが、IF の分布は分野によって異なることが指摘されている (Radicchi & Castellano, 2012; 小野寺, 2013)。また、先行研究では、IF が高いほどポリシーの掲載率や強度が高かった (Piwowar & Chapman, 2008; Stodden et al., 2013; Zenk-Möltgen & Lepthien, 2014)。このことから、分野間の比較を行う前に IF の高さとポリシーの掲載率や強度の関連を確認して、IF の影響が強い場合は調査対象誌の調整を行うこととした。

また、Stodden et al. (2013) の調査の結果や、先行研究の結果の比較 (2.5) によれば、調査年が新しいほどポリシーの掲載率が高かった。このことから、最近刊行された雑誌ほどポリシーの掲載率や強度が高い可能性があると推測される。そこで刊行開始年から調査時点までの年数を「雑誌年齢」と呼び、もし、雑誌年齢が若いほどポリシーの掲載率が高いならば、分野間で比較する際に雑誌年齢が影響しないよう、調査対象誌の調整を行うこととした。刊行年は、Ulrichweb.com の「開始年」で調べた。記載がない場合は雑誌のウェブサイトで確認した。

4.1.4 ポリシーの強度と公開先リポジトリ

予備調査と同様に、調査対象誌 (220 誌) のウェブサイトで投稿規定などを通読して、データ公開に関する記述やポリシーを調査した。投稿規定に記述が見当たらない場合は、“data”をキーワードとしてサイト内検索を行った。以下では、ポリシーの強度と公開先リポジトリの調査内容について述べる。

¹¹⁾ European Academy of Allergy and Clinical Immunology と International Union of Microbiological Societies の 2 機関は不明であったため、Ulrichsweb.com の出版国のままとした。

[1] ポリシーの強度

予備調査によって、ポリシーの強度は Stodden et al. (2013) のように「必須（査読・出版条件）」と「必須」を明確に区別できない場合があることがわかった（4.1.1）。そこで本調査では両者を区別せず、4段階の分類とした。すなわち、(1)データ公開を要求している場合や、査読・出版の条件としている場合は「必須」、(2)データ公開を奨励している場合は「推奨」、(3)データを受け付ける／公開について何らかの記述がある場合は「受諾」、(4)全く言及されていない場合は「なし」とした。Piwowar and Chapman (2008) や Zenk-Möltgen and Lepthien (2014) は「強い」、「弱い」、「なし」の3段階に分類していたが、共有を強く勧める「推奨」と、データを提出しても構わないとする「受諾」には明らかに差があると考えられることから4段階の分類とした。

表 4-5 に、ポリシーの強度と判断基準とした表現の例を示す。予備調査と同様に、ポリシーの対象となるデータの種類によって強度が異なる場合は、より高い方を採用することとした。

表 4-5 データ公開ポリシーの強度の分類

要求の強度	表現例
必須	require, must, should, condition of publication
推奨	encourage, recommend
受諾	accept, allow, might, possible to include
なし	(no mention)

[2] 公開先リポジトリ

予備調査では、データの公開先として GenBank などリポジトリ名を具体的に指定、または推奨している雑誌や、“パブリックリポジトリ”などと指示している雑誌がみられた。永続性のあるリポジトリはデータ公開の基盤とされている（Schofield, 2009）ことから、ポリシーを調べる際に、公開先リポジトリの例示の有無とリポジトリ名を調査した。

4.1.5 雑誌の特徴と仮説

本研究は、雑誌のポリシーを分野別に調査した上で比較を行うが、分野以外にもポリシーの掲載率や強度と関連がある特徴はあるのだろうか。また、どのような雑誌がデータ公開を要求しているのかを明らかにすることによって、雑誌がデータ公開を要求している理由を推測する手がかりにもなると考えられる。そこで先行研究と予備調査を参考に、データ公開を

要求する雑誌の特徴についての仮説を立てて、以下の3項目を調査した。

[1] 研究倫理と COI（利益相反）の開示に関する記述

データ公開の目的の1つに研究不正への対策がある（Arzberger, 2004; Guttmacher, Nabel, & Collins, 2009; Ioannidis, 2014）。予備調査で研究不正に関する記述がある雑誌がみられたことから「研究倫理や COI の開示に関する記述がある雑誌はポリシーの掲載率や要求の強度が高い」という仮説を立てて調査した。臨床試験報告に関する統合基準（Consolidated Standards of Reporting Trials, CONSORT）声明¹²⁾、出版倫理委員会（Committee on Publication Ethics, COPE）のガイドライン¹³⁾、世界医師会（World Medical Association, WMA）によるヘルシンキ宣言¹⁴⁾、医学雑誌編集者国際委員会（International Committee of Medical Journal Editors, ICMJE）による勧告¹⁵⁾などを採用しているかどうか調査した。なお、見出しを“Ethical Guidelines and Conflict of Interest”などとして、1箇所でも倫理と COI が併記されている場合もみられた。こうした場合はそれぞれカウントした。

[2] 出版者

先行研究では、学協会の雑誌の方が商業出版社の雑誌よりもポリシーの掲載率が高いという結果が示されていた（Piwowar & Chapman, 2008; Stodden et al., 2013）。そこで、「学協会の雑誌の方が商業出版社の雑誌よりもポリシーの掲載率や要求の強度が高い」という仮説を立てて出版者¹⁶⁾を調査した。

学協会と商業出版社の判別は、Ulrichsweb.com の「出版社および注文の詳細」画面と雑誌のウェブサイトで行った。なお、商業出版社のプラットフォームで公開されている雑誌であっても、出版主体が学協会である場合は学協会にカウントした。また、National Academy of Sciences のような学術機関も学協会とした。

¹²⁾ Consolidated Standards of Reporting Trials (CONSORT). <http://www.consort-statement.org>, (accessed 2018-10-31).

¹³⁾ “Guidelines,” The Committee on Publication Ethics (COPE). <http://publicationethics.org/resources/guidelines>, (accessed 2018-10-31).

¹⁴⁾ World Medical Association. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>, (accessed 2018-10-31).

¹⁵⁾ “Recommendations,” International Committee of Medical Journal Editors (ICMJE). <http://www.icmje.org/recommendations/>, (accessed 2018-10-31).

¹⁶⁾ 先行研究および本論文では、出版物を刊行する団体として「商業出版社」と「学協会」を取り上げるため、その総称として「出版者 (publisher)」という語を用いる。

[3] OA ステータス

オープンサイエンス政策では、公的資金による研究成果を市民に還元するという観点から、論文などの出版物に次いでデータへのオープンなアクセスを要求することが増えている（OECD, 2015; SPARC Europe, & Digital Curation Centre [DCC], 2018）。OA 誌はデータも論文と同様に公開するよう求めているのだろうか。先行研究では、OA 誌はポリシーの強度に影響を与えているとする結果（Piwowar & Chapman, 2008）と、OA 誌はポリシーの掲載率が比較的高い（Stodden et al., 2013）とする結果が示されていた。本研究では、「OA 誌はポリシーの掲載率や要求の強度が高い」という仮説を立てて調査した。

OA ステータスは急速に変化していると指摘されている（Björk, 2012）ため、既存のデータベースは使用せず、雑誌のウェブサイトで調査して「OA」、「制約つき OA」、「非 OA」に分類した。「制約つき OA」には、(1)購読者のみがアクセスできる非 OA 誌であるが、著者が追加の論文出版加工料（Article Processing Charge, APC）を支払うことによって特定の論文だけを OA にすることができるハイブリッド型の雑誌や、(2)著者や出版社が OA にするまでに一定の期間（エンバゴ）を設けている雑誌を含めた。

4.1.6 分析方法

4.1.3 で述べたように、IF と雑誌年齢がポリシーの強度に影響を与えていないかどうかを確認する際には、強度を 4（必須）、3（推奨）、2（受諾）、1（なし）の 4 段階の順序尺度として、Spearman の順位相関係数を調べた。有意水準 α は 5% として、有意な場合は相関係数 r と p 値を示した。また、雑誌のポリシーの掲載率と前章で調査したデータ公開率を分野ごとに比較する際には、Pearson の相関係数を調べることにした。その上で、 1σ （約 70%）の範囲、すなわち残差が 15% 以上または -15% 以下の場合は外れ値とみなし、要求の状況に対して公開率が高い分野や低い分野であるとして特定することとした。ソフトウェアは SPSS Version 24 を用いた。

4.2 結果

4.2.1 分野別データ公開ポリシー

表 4-6 に、リポジトリにデータを公開して論文に識別子を記すポリシー（Repository Policy, 以下「RP」）と、補足資料にデータを掲載するポリシー（Supplemental Materials' Policy, 以下

4 章 データ公開の外的要因と日本の研究者の状況

「SP」)の掲載状況を示す。調査対象とした220誌のうち、RPは131誌(59.5%)¹⁷⁾、SP¹⁸⁾は197誌(89.5%)が掲載していた。両方を掲載している雑誌は124誌(56.4%)、いずれも掲載していない雑誌は16誌(7.3%)であり、204誌(92.7%)がRPまたはSPを掲載していた。

表 4-6 データ公開ポリシーの掲載状況

RP \ SP	あり		なし		合計	
	雑誌数	比率	雑誌数	比率	雑誌数	比率
あり	124	56.4%	7	3.2%	131	59.5%
なし	73	33.2%	16	7.3%	89	40.5%
合計	197	89.5%	23	10.5%	220	100.0%

表 4-7 に、分野別の集計結果とデータの公開先としてリポジトリを指定していた雑誌数を示す。分野の排列は、RP の掲載誌数が多い順として、同数の場合はポリシーの強度がより高い雑誌が多い順とした。

¹⁷⁾ *The Journal of Physical Chemistry Part C* は、結晶構造や表面構造データは補足資料 (Supporting Information) として CIF 形式で提出することを「必須」としている。補足資料のポリシーは別にあり、データの「受諾」ととどまるため両者は区別されるが、登録先が CSD のような外部リポジトリではなく補足資料であるため、リポジトリによるデータ公開ポリシーは「なし」とした。“Guidelines for Structural Data”. *The Journal of Physical Chemistry Part C: Nanomaterials, Interfaces and Hard Matter*. http://pubs.acs.org/paragonplus/submission/jpchax/jpchax_authguide.pdf, (accessed 2014-5-8).

¹⁸⁾ 調査対象誌の「補足資料」の表現には、“Supplemental Materials”, “Supplemental Information”, “Supporting Data”, “Additional File”などがあった。

表 4-7 分野別学術雑誌のデータ公開ポリシー (n=220)

分野	リポジトリ (RP)					補足資料 (SP)					公開先
	あり	必須	推奨	受諾	なし	あり	必須	推奨	受諾	なし	
A 分子生物学	10	10	0	0	0	10	3	0	7	0	10
生物学	10	9	1	0	0	10	5	1	4	0	10
臨床医学	10	9	1	0	0	10	2	0	8	0	10
微生物学	9	8	0	1	1	10	1	2	7	0	9
材料科学	8	8	0	0	2	10	3	1	6	0	8
植物動物学	8	8	0	0	2	9	2	0	7	1	8
化学	8	7	1	0	2	10	3	4	3	0	8
神経科学	8	7	1	0	2	9	1	1	7	1	6
環境学	8	5	3	0	2	10	2	2	6	0	8
免疫学	8	5	3	0	2	10	1	1	8	0	8
B 複合領域	7	7	0	0	3	9	4	1	4	1	7
農業科学	7	3	4	0	3	9	1	1	7	1	6
地球科学	7	2	5	0	3	10	1	1	8	0	7
薬理学	6	4	2	0	4	10	0	2	8	0	6
宇宙科学	6	1	5	0	4	9	0	5	4	1	6
社会科学	4	2	2	0	6	7	0	2	5	3	2
C 物理学	2	2	0	0	8	10	2	2	6	0	2
CS	2	2	0	0	8	9	0	1	8	1	2
精神医学	2	1	1	0	8	8	0	1	7	2	1
数学	1	0	1	0	9	3	0	1	2	7	0
工学	0	0	0	0	10	9	0	0	9	1	0
経済学	0	0	0	0	10	6	0	0	6	4	0
合計	131	100	30	1	89	197	31	29	137	23	124
(%)	59.5	45.5	13.6	0.5	40.5	89.5	14.1	13.2	62.3	10.5	56.4

「公開先」=データの公開先リポジトリを指定していた雑誌数

ポリシーの掲載率や強度は分野による差がみられたため、これらを勘案してグループに分けて分析することとした。具体的には、RPの掲載誌数が10～8の10分野を「グループA」、7～4の6分野を「グループB」、2～0の6分野を「グループC」とした。

ここで、IF や雑誌年齢が分野別の結果に影響を与えているかどうかを確認した。すなわち IF が高い、あるいは雑誌年齢が若い雑誌を多く含む分野が、ポリシーの掲載率や強度が高くなっていないかどうかを確認した。

表 4-8 に、分野別 IF の中央値、最小値、最大値、標準偏差を示す。全 220 誌の IF の中央値は 6.4 であり、臨床医学 (24.7) や分子生物学・遺伝学 (14.5) は高く¹⁹⁾、数学 (3.5)、農業科学 (3.6) は低い傾向にあった。また、複合領域、臨床医学、材料科学の雑誌は分野内での IF のばらつきが比較的大きかった。

全 220 誌の IF の高さ と RP の強度の関連を Spearman の相関係数を用いて確認したところ、中程度の相関が認められた ($r=0.520$, $p<0.01$)。同様に、IF の高さ と SP の強度の相関を確認したところ、弱い相関が認められた ($r=0.417$, $p<0.01$)。また、分野の影響を排して IF とポリシー強度の関連をみるため、分野ごとに IF が高い順に 10 から 1 の順位をつけて Spearman の相関分析を行った。IF の順位と RP の強度は $r=0.178$, $p<0.01$, IF の順位と SP の強度は $r=0.240$, $p<0.01$ であり、ほとんど相関は認められなかった (集計結果は付表 52 に示す)。以上の結果から IF の違いによる影響はそれほど大きくないと判断した。

¹⁹⁾ *New England Journal of Medicine* (51.7), *Lancet* (39.1), *Nature* グループの雑誌などの値が高かった。

表 4-8 分野別 IF (n=220)

	分野	中央値	最小値	最大値	標準偏差
A	1 分子生物学	14.5	11.2	35.2	8.24
	2 生物学	12.4	9.5	32.4	7.50
	3 臨床医学	24.7	17.2	51.7	10.04
	4 微生物学	5.4	4.6	12.6	2.46
	5 材料科学	9.4	6.1	35.7	10.01
	6 植物動物学	6.3	5.3	9.3	1.05
	7 化学	9.5	6.2	21.8	4.60
	8 神経科学	9.8	6.9	15.8	3.03
	9 環境学	7.8	6.3	17.9	3.73
	10 免疫学	6.7	5.5	26.2	6.77
	11 複合領域	3.9	0.9	38.6	12.74
B	12 農業科学	3.6	3.4	4.3	0.33
	13 地球科学	5.1	4.4	12.4	2.31
	14 薬理学	5.1	4.6	9.2	1.60
	15 宇宙科学	5.3	3.3	16.2	3.48
	16 社会科学	4.2	4.1	5.8	0.51
C	17 物理学	7.2	4.7	27.3	7.13
	18 CS	4.4	3.6	9.3	1.58
	19 精神医学	6.8	5.6	14.7	3.12
	20 数学	3.5	2.5	6.0	1.03
	21 工学	4.8	4.4	7.7	1.06
	22 経済学	4.9	4.1	6.7	0.94
全体		6.4	0.9	51.7	7.80

表 4-9 に分野別雑誌年齢の中央値、最小値、最大値、標準偏差を示す²⁰⁾。調査を実施した 2014 年時点での全 220 誌の雑誌年齢の中央値は 27.0 年であった。微生物学 (9.5 年) と物理学 (9.5 年) は低く、複合領域 (73.0 年) や宇宙科学 (72.0 年) は高いなど、分野による差がみられた。また、臨床医学と宇宙科学は、分野内でのばらつきが大きかった。刊行開始年が最も古いタイトルは 1823 年の *Lancet*、最新は 2011 年の *Advanced Energy Materials* ほか 3 誌であった。付表 53 に刊行開始年を 10 年ごとに集計した結果を示す。

²⁰⁾ Ulrichsweb.com に記載がなかった *The Astronomical Journal* のみ雑誌のウェブサイトで確認した。

4 章 データ公開の外的要因と日本の研究者の状況

雑誌年齢とポリシーの強度に関連がみられるかどうかを確認するために Spearman の相関分析を行ったところ、わずかに負の相関が認められた (RP が $r=-0.197$, $p<0.01$, SP が $r=-0.238$, $p<0.01$)。つまり、新しく創刊された雑誌の方が、ややポリシーの強度が高い程度であった。また、分野の影響を排して雑誌年齢とポリシー強度の関連をみるため、分野ごとに雑誌年齢が低い順に 1 から 10 の順位をつけて Spearman の相関分析を行ったが、雑誌年齢と RP の強度、雑誌年齢と SP の強度はいずれも相関が認められなかった (集計結果は付表 54 に示す)。以上の結果から雑誌年齢の違いによる影響はほとんど認められないと判断した。

表 4-9 分野別雑誌年齢 (n=220)

	分野	中央値	最小値	最大値	標準偏差
A	1 分子生物学	19.5	9	66	16.54
	2 生物学	16.0	9	38	10.29
	3 臨床医学	25.0	7	191	70.42
	4 微生物学	9.5	4	47	12.36
	5 材料科学	17.5	3	34	9.89
	6 植物動物学	29.5	11	112	31.96
	7 化学	10.0	4	171	58.19
	8 神経科学	35.0	16	136	33.51
	9 環境学	20.5	3	83	24.47
	10 免疫学	27.0	6	110	36.27
	11 複合領域	73.0	3	145	54.62
B	12 農業科学	31.5	6	86	23.21
	13 地球科学	33.5	6	94	26.09
	14 薬理学	30.5	7	84	25.70
	15 宇宙科学	72.0	4	187	61.39
	16 社会科学	38.5	22	91	24.01
C	17 物理学	9.5	3	121	44.23
	18 CS	19.0	7	46	12.27
	19 精神医学	56.5	8	170	52.58
	20 数学	65.0	5	140	44.98
	21 工学	32.0	17	101	23.87
	22 経済学	45.0	9	128	30.14
	全体	27.0	3	191	41.56

IF と雑誌年齢の分野による差はそれほど大きくなかったことから、調査対象誌 220 誌の調

整（入れ替え）は行わずに分野間比較を行うこととした。

図 4-2 に分野別 RP の強度分布を示す。RP を掲載している 131 誌（全体の 59.5%，数値は表 4-7 参照）の強度を確認すると、「必須」（45.5%，図ではオレンジ色の部分）と「推奨」（13.6%，薄オレンジ色）の比率が高く、「受諾」（0.5%，水色）は微生物学の 1 誌のみであった。つまり、RP を掲載していた雑誌は要求の強度が比較的高かった。また、分野による差が大きく、分子生物学・遺伝学，生物学・生化学，臨床医学は 10 誌全てに掲載されていたのに対して，経済学・経営学と工学は全く掲載されていなかった。

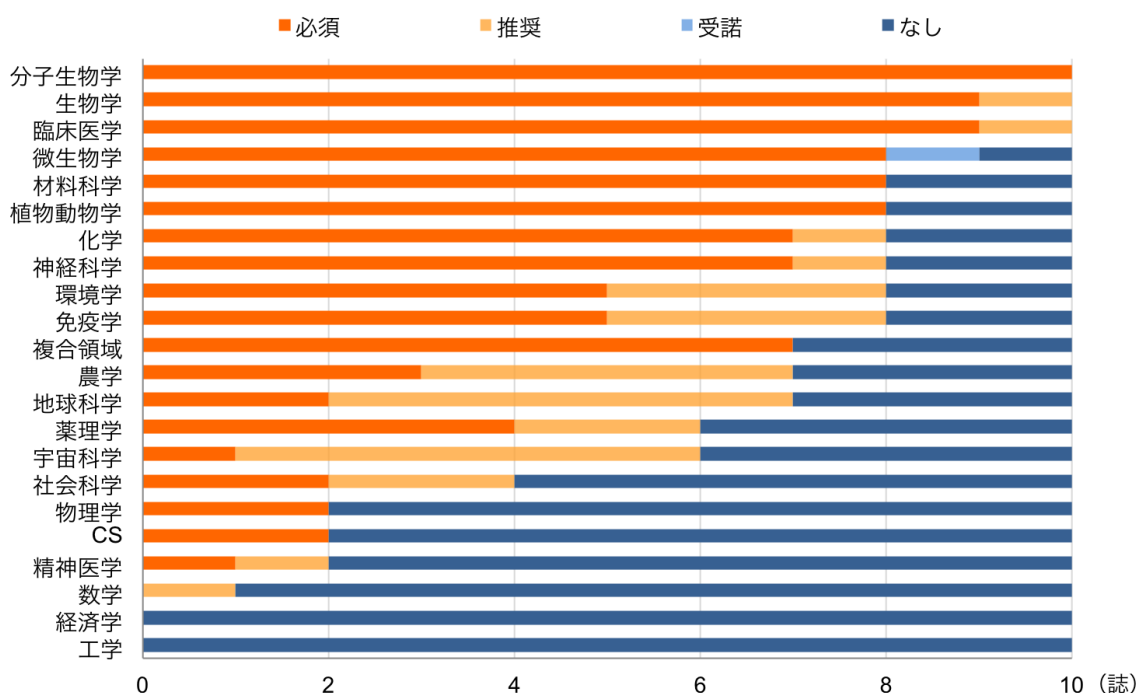


図 4-2 分野別データ公開ポリシーの強度分布（22 分野／各 10 誌）

表 4-7 の「公開先」に示した通り，124 誌（56.4%）が公開先のリポジトリを指示していた。うち，121 誌はデータの公開先として具体的なリポジトリ名を例示していた。また，こうしたリポジトリ名を例示した上で，*Molecular Ecology Resources*（環境・生態学）は機関リポジトリか助成機関のリポジトリで²¹⁾，*Genome Research*（分子生物学・遺伝学）は著者のウェブサイトでもデータを公開することも認めていた（各 1 誌のみ）。残る 3 誌は既存のディレクトリからリポジトリを選ぶよう指示していた。*Retrovirology* は DataCite のリポジトリリストから²²⁾，

²¹⁾ GenBank や Gene Expression Omnibus などと並んで“your own institutional or funder repository”と記されている。“Policy on Data Archiving”. *Molecular Ecology Resources*.

[http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1755-0998/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1755-0998/homepage/ForAuthors.html), (accessed 2018-10-31).

²²⁾ “Availability of supporting data”. *Retrovirology*.

Atmospheric Chemistry and Physics と *Geoscientific Model Development* の 2 誌は re3data.org²³⁾あるいは Databib²⁴⁾から²⁵⁾適切なリポジトリを選ぶよう指示していた。

表 4-10 に、121 誌に例示されていたデータの公開先リポジトリの集計結果から、上位 10 件を示す。GenBank や Gen Expression Omnibus (GEO) など、生物・医学分野のリポジトリが多数例示されていた。

表 4-10 データの公開先として例示されたりポジットリ (上位 10 件)

リポジトリ名	雑誌数	開始年	主なデータ
GenBank [†]	79	1983	塩基配列
GEO (Gen Expression Omnibus)	64	2002	遺伝子発現
ArrayExpress	59	1994	遺伝子発現
DDBJ (DNA Databank of Japan) [†]	59	1986	塩基配列
EMBL (European Molecular Biology Laboratory) [†]	56	1980	塩基配列
PDB (Protein Data Bank)	50	1971	蛋白質
UniProt/Swiss-Prot/PIR	46	2002	蛋白質
CSD (Cambridge Structural Database)	40	1987	結晶構造
PANGAEA	22	1996	地球科学・環境科学
Dryad	19	2008	自然科学・医学

[†]GenBank, DDBJ, EMBL は INSDC (統一フォーマットの国際塩基配列データベース) で相互運用されているが、いずれか 1 つ、2 つ、あるいは全てを例示する雑誌がみられたため、それぞれ集計した。

図 4-3 に分野別 SP の強度分布を示す。SP は約 9 割 (89.5%) の雑誌が掲載しており、かつ、全ての分野で掲載されていた。ただし、強度が低い「受諾」が 137 誌 (62.3%) であり (表 4-7)、約 6 割の雑誌は提出を可能とするという程度にとどまっていた (図 4-3, 水色部分)²⁶⁾。

<http://www.retrovirology.com/authors/instructions/research#formatting-supporting-data>, (accessed 2014-5-8).

²³⁾ re3data.org (Registry of Research Data Repositories) は、2012 年にドイツ研究振興協会 (Deutsche Forschungsgemeinschaft, DFG) の助成を受けて開始されたデータリポジトリのレジストリであり、EC の “Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020”において、適切なデータリポジトリの選択に用いるよう推奨されている。 <https://www.re3data.org>, (accessed 2018-10-31).

²⁴⁾ データリポジトリのレジストリ。2015 年に re3data.org とともに DataCite に統合された。2018 年 10 月現在、レジストリの名称は上記の re3data.org が使われている。

²⁵⁾ “Data Policy”. *Atmospheric Chemistry and Physics*. http://www.atmospheric-chemistry-and-physics.net/general_information/data_policy.html, (accessed 2014-5-8).

²⁶⁾ 補足資料を査読対象とすることを明示していたのは、54 誌 (全体の 24.5%) であった。記述は雑誌によってさまざまであり、査読者が公開するかどうかを決める (*Journal of Virology*)、査読の上で著者が公開するかどうかを決める (*Journal of Nutrition*)、査読するが著者が責任をもつ (*JAMA*) といった例がみられた。

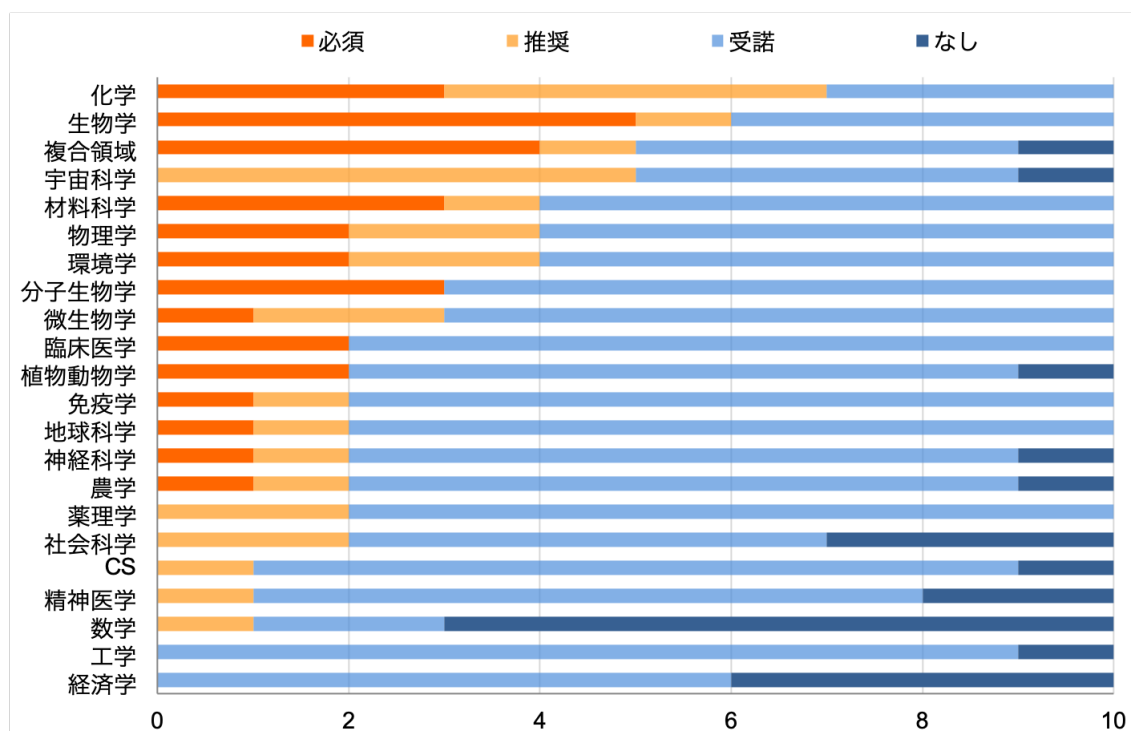


図 4-3 分野別補足資料ポリシーの強度分布 (22 分野／各 10 誌)

補足資料の公開は出版者にとって負担となるため²⁷⁾、多くの雑誌はデータ量やファイル数の制限、課金といった対策を講じていた。*The Journal of Neuroscience* は一論文あたりのデータ量が急増したことから (Maunsell, 2010)、2010 年 11 月以降、補足資料を廃止した²⁸⁾。Cell Press は補足資料の有用性に触れつつも、長大化しているために著者、査読者、読者に負担をかけていると指摘した上で、補足資料を構造化することや 50MB 以下に制限することを求めている²⁹⁾。*Molecular Biology and Evolution*³⁰⁾は、ファイル形式を指定した上で、5 ファイル、各 2MB まで、*Journal of Immunology*³¹⁾は、ビデオは 30 秒以内、10MB まで、図表は 1 件あたり \$70 としている。*Science*³²⁾や *Scientific Report*³³⁾などは、適切な外部リポジトリがない場合限り、掲載を認めていた。なお、調査対象誌のうち *Systematic Biology* だけは、Dryad にデータ

²⁷⁾ 非 OA 誌や非 OA 論文であっても、補足資料のみ公開している雑誌もみられた。

²⁸⁾ ただし、外部のウェブサイトに登録・保存したデータへのリンクを論文に含めることは可能である。

“Supplemental Material”. *The Journal of Neuroscience*. http://www.jneurosci.org/site/misc/ifa_supplemental.xhtml, (accessed 2014-5-8).

²⁹⁾ “Supplemental Information Guidelines”. Cell Press. <http://www.cell.com/supplemental-information>, (accessed 2014-5-8).

³⁰⁾ “Supplementary Information”. *Molecular Biology and Evolution*. http://www.oxfordjournals.org/our_journals/molbev/supplementary_information.html, (accessed 2014-5-8).

³¹⁾ “Supplemental Materials”. *Journal of Immunology*. <http://www.jimmunol.org/site/misc/authorinstructions.xhtml#supplemental>, (accessed 2014-5-8).

³²⁾ “Supplementary Materials”. *Science*. http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail, (accessed 2014-5-8).

³³⁾ “Availability of materials and data”. *Scientific Reports*. <http://www.nature.com/srep/policies/index.html#availability>, (accessed 2014-5-8).

を登録した上で、補足資料に“Data available from the Dryad Digital Repository”の見出しとともにデジタルオブジェクト識別子 (Digital Object Identifier, DOI) を記述するよう指示していた³⁴⁾。

以下では、RP の掲載誌数が 10~8 の 10 分野を「グループ A」、7~4 の 6 分野を「グループ B」、2~0 の 6 分野を「グループ C」として、それぞれの調査結果を述べる。

[1] ポリシーの掲載率と強度が高い分野 (グループ A)

グループ A には、分子生物学・遺伝学や臨床医学など広義の生物・医学分野、および材料科学、化学、環境・生態学が含まれていた。グループ A の 11 分野各 10 誌のいずれかは、表 4-10 に示した GenBank や Gen Expression Omnibus (GEO) など生物データのリポジトリを必ず例示していた。また、材料科学と化学は Cambridge Structural Database (CSD) などの結晶構造のリポジトリを、環境・生態学は PANGAEA や Dryad を例示していた。

グループ A の 10 分野各 10 誌は、RP か SP のどちらか一方、あるいは両方を掲載していた。つまり、RP も SP も掲載していない雑誌はなかった。

[2] ポリシーの掲載率と強度が中程度の分野 (グループ B)

公開データ (一次データ) の二次分析が盛んな地球科学、宇宙科学、社会科学は、ポリシーの掲載率や強度が中程度であった。これらの分野では、研究対象とした一次データを特定するためのリンクの記述などを求めている雑誌が多かった。たとえば宇宙科学の *The Astrophysical Journal* など 4 誌は、NASA Archive にデータがあればリンクさせるよう指示しており³⁵⁾、社会科学の *Health Reports* は社会統計などの元データへの記述を含まなければならないとしていた³⁶⁾。

複合領域、農業科学、薬理学・毒性学のうち、RP を掲載している雑誌は、いずれも表 4-10 にみられる生物データのリポジトリを挙げていた。ただし、農業科学は「必須」が 3 誌、「推奨」が 4 誌、薬理学・毒性学は「必須」が 4 誌、「推奨」が 2 誌であり、他の生物学分野と比較して掲載率や強度が低かった。

グループ B で RP と SP のいずれもなかった雑誌は、農業科学、宇宙科学、社会科学の各 1

³⁴⁾ 同誌はリポジトリによるデータ共有も「必須」としており、登録先として GenBank や EMBLなどを挙げている。“Data and Supplementary Materials”. *Systematic Biology*. http://www.oxfordjournals.org/our_journals/sysbio/for_authors/ms_preparation.html, (accessed 2014-5-8).

³⁵⁾ Institute of Physics Publishing の *Astrophysical Journal*, *Astrophysical Journal Supplement Series*, *Astrophysical Journal Letters*, *Astronomical Journal* <http://aas.org/authors/manuscript-preparation-aj-apj-author-instructions>, (accessed 2014-5-8).

³⁶⁾ “Methods”. *Health Reports*. <http://www.statcan.gc.ca/pub/82-003-x/4060781-eng.htm#n3>, (accessed 2014-5-8).

誌, 合計 3 誌であった。つまり, ほとんどの雑誌に RP か SP のどちらか一方, あるいは両方が掲載されていた。

[3] ポリシーの掲載率や強度が低い分野（グループ C）

物理学は, RP は「必須」2 誌のみであった。しかし, 「なし」の雑誌であってもデータに関する記述はみられた。たとえば, *The Journal of Physical Chemistry Part C: Nanomaterials, Interfaces and Hard Matter* は構造データを補足情報 (Supplemental Information) として提供するように要求しており, *European Physical Journal C* にはデータの改竄について言及していた。一方, 数学は「推奨」1 誌のみであり, 当該誌の *Database: The Journal of Biological Databases and Curation* はデータベースと情報源を少なくとも 2 年間アクセス可能にするよう推奨していた³⁷⁾。ただし, 公開先を指示しておらず期間も限られていた。数学は SP も 7 誌が「なし」であり, 22 分野で最も掲載率が低かった。

コンピュータサイエンス, 工学, 経済学・経営学のうち RP を掲載していたのはコンピュータサイエンスの 2 誌のみであったが, 研究に用いたデータに関する記述は各分野でみられた。たとえば, IEEE の雑誌はコンピュータサイエンスに 4 誌, 工学に 5 誌含まれているが, 提出するデータの形式について Frequently Asked Questions (FAQ) で言及していた³⁸⁾。また, 経済学・経営学の実証研究に関する論文を掲載している *Academy of Management Journal* は, 論文に使われたデータが過去にも使われたかどうかを明らかにするよう求めている (Colquitt, 2013)。つまり, この 3 分野は数学とは異なり, データを用いた研究が行われているにもかかわらず, ポリシーの掲載率が低かった。

精神医学・心理学の RP は, 「必須」と「推奨」が各 1 誌ずつ掲載されているのみであった。*British Journal of Psychiatry* はデータの公開を“強く推奨”していたものの, 公開先は指定していなかった³⁹⁾。

グループ C の 6 分野 60 誌のうち, RP と SP のいずれも掲載していなかったのは, 数学が 6 誌, 経済学・経営学が 4 誌, 精神医学・心理学が 2 誌, 工学が 1 誌, 合計 13 誌であった。つまりグループ C のうち, 約 2 割の雑誌は全くポリシーを掲載していなかった。

4.2.2 雑誌の特徴：倫理と COI の開示, 出版者, OA ステータス

表 4-11 に, 調査対象誌の特徴として, 倫理や COI の記述がある雑誌数, 出版者が学協会

³⁷⁾ “Availability of databases”. Database: The Journal of Biological Databases and Curation. http://www.oxfordjournals.org/our_journals/databa/for_authors/general.html, (accessed 2014-5-8).

³⁸⁾ “Author Frequently Asked Questions”. IEEE: Institute of Electrical and Electronics Engineers. http://www.ieee.org/publications_standards/publications/authors/publications_faq.html, (accessed 2014-5-8).

³⁹⁾ “Access to data”. British Journal of Psychiatry. <http://bjp.rcpsych.org/site/misc/ifora.xhtml#General>, (accessed 2014-5-8).

である雑誌数, OA と制約つき OA の雑誌数を示す。

表 4-11 分野別調査対象誌の特徴 (n=220)

	分野	雑誌の特徴			
		倫理	COI	学協会	OA [†]
A	1 分子生物学	7	8	2	8
	2 生物学	9	10	2	5
	3 臨床医学	4	9	4	6
	4 微生物学	8	8	3	7
	5 材料科学	8	5	2	7
	6 植物動物学	4	4	8	7
	7 化学	8	8	8	4
	8 神経科学	6	8	4	8
	9 環境学	8	8	5	6
	10 免疫学	9	10	4	8
B	11 複合領域	6	9	6	6
	12 農業科学	6	6	5	10
	13 地球科学	6	4	7	8
	14 薬理学	8	9	5	6
	15 宇宙科学	3	1	8	4
	16 社会科学	4	6	7	7
C	17 物理学	7	7	6	7
	18 CS	3	4	6	8
	19 精神医学	6	8	9	6
	20 数学	3	4	8	5
	21 工学	3	3	5	4
	22 経済学	3	4	7	3
合計		129	143	121	140

†OA は制約つき OA 誌を含む

表 4-11 に示したように雑誌の特徴は分野による差がみられた。有意水準 5% で二項検定を行ったところ、倫理は 2 分野、COI は 6 分野、学協会は 5 分野、OA は 1 分野が母集団から外れていた。つまり、OA 以外は分野によって特徴の分布が大きく異なるといえる。そこで倫理、COI、学協会（出版者）は全 220 誌の結果を示した後に、分野による交絡を制御するため生物・医学と自然科学の「分野グループ（表 4-12）」に分けた結果を示す。

表 4-12 分野グループ

生物・医学分野	自然科学分野	その他
分子生物学・遺伝学, 生物学・生化学, 臨床医学, 微生物学, 植物・動物学, 神経科学・行動科学, 免疫学, 農業科学, 薬理学・毒性学, 精神医学・心理学	材料科学, 化学, 環境・生態学, 地球科学, 宇宙科学, 物理学, コンピュータサイエンス, 数学, 工学	複合領域, 社会科学, 経済学・経営学
10 分野	9 分野	3 分野

[1] 倫理と COI の開示

表 4-13 に倫理と COI に関する記述の有無とデータ公開ポリシーの有無の集計結果を示す。倫理に関する記述を含む雑誌は 129 誌（全体の 58.6%）であった。その主な内容はデータの改竄や捏造、オーサーシップといった出版倫理に関する規定や、ヒトや動物を対象とした研究に関する規定であった。COI の開示に関する記述を含む雑誌は 143 誌（全体の 65.0%）であり、資金提供や共同研究者など研究に影響を与えると考えられる利害関係の事実を開示するよう求めている。たとえば *Tobacco Control* は、タバコ会社やタバコ業界団体の資金提供を受けた論文は認めていなかった。国際的な指針に準拠している例も多く、CONSORT 声明は 38 誌、COPE ガイドラインは 27 誌、ヘルシンキ宣言は 25 誌、ICMJE による勧告は 18 誌で採用あるいは言及されていた。

表 4-13 倫理・COI に関する記述の有無とデータ公開ポリシーの有無

記述	雑誌数	データ公開ポリシー	
		あり	なし
倫理	有	129	3
		97.7%	2.3%
	無	91	13
		85.7%	14.3%
COI	有	143	4
		97.2%	2.8%
	無	77	12
		84.4%	15.6%
合計	220	204	16
		92.7%	7.3%

ポリシーが掲載されていた雑誌 (RP と SP のいずれか、または両方が掲載されていた雑誌) は、倫理の記述有りの雑誌 (以下、「有り」) は 129 誌中 126 誌で 97.7%、倫理の記述無しの雑誌 (以下、ポリシーがないことの「なし」と区別して「無し」と記す) は 91 誌中 78 誌で 85.7%であり、倫理の記述有りの方が掲載率が高かった。COI の記述有りの雑誌は 143 誌中 139 誌で 97.2%、COI の記述無しは 77 誌中 65 誌で 84.4%であり、倫理と同様に COI の記述有りの方が掲載率が高かった。

表 4-14 に倫理と COI に関する記述の有無ごとの RP と SP の掲載率を示す。

倫理の記述有りの雑誌は RP の掲載率が 70.5%、無しは 44.0%であり、有りの雑誌の方が高かった。倫理の記述有りの雑誌の SP の掲載率は 94.6%、無しは 82.4%であり、同じく有りの雑誌の方が高かった。

COI の記述有りの雑誌は RP の掲載率が 69.2%、無しは 41.6%であり、有りの雑誌の方が高かった。COI の記述有りの雑誌の SP の掲載率は 94.4%、無しは 80.5%であり、同じく有りの雑誌の方が高かった。

表 4-14 倫理・COI に関する記述の有無とデータ公開ポリシーの掲載率

記述	雑誌数	リポジトリ (RP)		補足資料 (SP)	
		あり	なし	あり	なし
倫理	有	129	91	38	122
			70.5%	29.5%	94.6%
	無	91	40	51	75
			44.0%	56.0%	82.4%
COI	有	143	99	44	135
			69.2%	30.8%	94.4%
	無	77	32	45	62
			41.6%	58.4%	80.5%
合計		220	131	89	197
			59.5%	40.5%	89.5%

表 4-15 に、倫理と COI の開示に関する記述の有無とポリシーの強度を示す。

倫理の記述有りの雑誌は、RP が「必須」の比率 (52.7%) が無しの雑誌 (35.2%) よりも高く、「必須」と「推奨」の合計も有りは 69.8%、無しは 44.0%でやはり高かった。倫理有りの雑誌は SP が「必須」の比率 (19.4%) も無しの雑誌 (6.6%) よりも高く、「必須」と「推奨」の合計も有りの雑誌は 34.1%、無しの雑誌は 17.6%で同じく高かった。

COI の記述有りの雑誌は、RP が「必須」の比率（54.5%）が無しの雑誌（28.6%）よりも高く、「必須」と「推奨」の合計も有りの雑誌は 68.5%、無しの雑誌は 41.6%でやはり高かった。COI 有りの雑誌は SP が「必須」の比率（21.0%）が無しの雑誌（1.3%）よりも高く、「必須」と「推奨」の合計も有りの雑誌は 32.2%、無しの雑誌は 18.2%で同じく高かった。

表 4-15 倫理・COI に関する記述の有無とデータ公開ポリシーの強度

記述	雑誌数	リポジトリ (RP)				補足資料 (SP)				
		必須	推奨	受諾	なし	必須	推奨	受諾	なし	
倫理	有	129	68	22	1	38	25	19	78	7
		52.7%	17.1%	0.8%	29.5%	19.4%	14.7%	60.5%	5.4%	
	無	91	32	8	0	51	6	10	59	16
		35.2%	8.8%	0.0%	56.0%	6.6%	11.0%	64.8%	17.6%	
COI	有	143	78	20	1	44	30	16	89	8
		54.5%	14.0%	0.7%	30.8%	21.0%	11.2%	62.2%	5.6%	
	無	77	22	10	0	45	1	13	48	15
		28.6%	13.0%	0.0%	58.4%	1.3%	16.9%	62.3%	19.5%	
合計		220	100	30	1	89	31	29	137	23
		45.5%	13.6%	0.5%	40.5%	14.1%	13.2%	62.3%	10.5%	

表 4-16 に分野グループ別の結果を示す。倫理の記述有りは生物・医学が 100 誌中 67 誌（67.0%）、自然科学が 90 誌中 49 誌（54.4%）であった。COI の記述有りは生物・医学が 100 誌中 80 誌（80.0%）、自然科学は 90 誌中 44 誌（48.9%）であり、いずれも生物・医学の方が掲載率が高かった。

表 4-16 倫理・COI に関する記述の有無とデータ公開ポリシー（分野グループ別）

分野 グループ	雑誌数	リポジトリ (RP)				補足資料 (SP)					
		必須	推奨	受諾	なし	必須	推奨	受諾	なし		
倫理	生物・医学	有	67	42	12	1	12	13	8	43	3
				62.7%	17.9%	1.5%	17.9%	19.4%	11.9%	64.2%	4.5%
		無	33	22	1	0	10	3	1	27	2
				66.7%	3.0%	0.0%	30.3%	9.1%	3.0%	81.8%	6.1%
	自然科学	有	49	22	8	0	19	10	10	28	1
				44.9%	16.3%	0.0%	38.8%	20.4%	20.4%	57.1%	2.0%
		無	41	5	7	0	29	1	7	24	9
				12.2%	17.1%	0.0%	70.7%	2.4%	17.1%	58.5%	22.0%
COI	生物・医学	有	80	54	11	1	14	16	6	56	2
				67.5%	13.8%	1.3%	17.5%	20.0%	7.5%	70.0%	2.5%
		無	20	10	2	0	8	0	3	14	3
				50.0%	10.0%	0.0%	40.0%	0.0%	15.0%	70.0%	15.0%
	自然科学	有	44	17	7	0	20	10	8	24	2
				38.6%	15.9%	0.0%	45.5%	22.7%	18.2%	54.5%	4.5%
		無	46	10	8	0	28	1	9	28	8
				21.7%	17.4%	0.0%	60.9%	2.2%	19.6%	60.9%	17.4%
合計	190	91	28	1	70	27	26	122	15		
			47.9%	14.7%	0.5%	36.8%	14.2%	13.7%	64.2%	7.9%	

倫理の記述について、RP、SP の掲載率は、生物・医学、自然科学ともに記述有りの方が無しよりも高かった。生物・医学で、RP の掲載率（全体から「(ポリシー) なし」を引いた値）は、倫理の記述有りの雑誌は 82.1%、無しの雑誌は 69.7%であった。SP の掲載率は、倫理の記述有りの雑誌は 95.5%、無しの雑誌は 93.9%であった。RP の強度は、倫理の記述有りの雑誌は「必須」が 62.7%、無しの雑誌は 66.7%であり、無しの方が高かった。「必須」と「推奨」の合計は、有りの雑誌は 80.6%、無しの雑誌は 69.7%であった。SP の強度は、倫理の記述有りの雑誌は「必須」が 19.4%、無しの雑誌は 9.1%であり、「必須」と「推奨」の合計は倫理の記述有りの雑誌は 31.3%、無しの雑誌は 12.1%であった。

自然科学で、倫理の記述有りの雑誌の RP の掲載率は 61.2%、無しの雑誌は 29.3%であった。SP の掲載率は、倫理の記述有りの雑誌は 98.0%、無しの雑誌は 78.0%であった。RP の強度は、倫理の記述有りの雑誌は「必須」が 44.9%、無しの雑誌は 12.2%であり、「必須」と「推奨」の合計が有りの雑誌は 61.2%、無しの雑誌は 29.3%であった。SP の強度は、倫理の記述

有りの雑誌は「必須」が20.4%、無しの雑誌は2.4%であり、「必須」と「推奨」の合計が倫理の記述有りの雑誌は40.8%、無しの雑誌は19.5%であった。

COIの記述について、RP、SPによるポリシーの掲載率は、生物・医学、自然科学ともに記述有りの雑誌の方が無しの雑誌よりも高かった。生物・医学で、RPの掲載率（全体から「(ポリシー) なし」を引いた値）は、COIの記述有りの雑誌は82.5%、無しの雑誌は60.0%で差があった。SPの掲載率は、COIの記述有りの雑誌は97.5%、無しの雑誌は85.0%で差がみられた。SPの強度については、COIの記述有りの雑誌は「必須」が20.0%、無しの雑誌は0.0%であり、「必須」と「推奨」の合計はCOIの記述有りの雑誌は27.5%、無しの雑誌は15.0%であった。

自然科学で、COIの記述有りの雑誌のRPの掲載率は54.5%、無しの雑誌は39.1%であった。SPの掲載率は、COIの記述有りの雑誌は95.5%、無しの雑誌は82.6%であった。SPの強度については、COIの記述有りの雑誌は「必須」が22.7%、無しの雑誌は2.2%であり、「必須」と「推奨」の合計はCOIの記述有りの雑誌は40.9%、無しの雑誌は21.8%であった。

[2] 出版者

表4-17に、学協会と商業出版社別のデータ公開ポリシーの有無の集計結果を示す。調査対象の220誌のうち、学協会の雑誌は121誌（55.0%）、商業出版社の雑誌は99誌（45.0%）であった。ポリシーが掲載されていた雑誌（RPとSPのいずれか、または両方が掲載されていた雑誌）は、学協会が121誌中109誌で90.1%、商業出版社が99誌中95誌で96.0%であり、商業出版社の方が比率が高かった。

表 4-17 学協会・商業出版社別データ公開ポリシーの有無

出版者	雑誌数	データ公開ポリシー	
		あり	なし
学協会	121	109	12
		90.1%	9.9%
商業出版社	99	95	4
		96.0%	4.0%
合計	220	204	16
		92.7%	7.3%

表4-18に学協会と商業出版社別のRPとSPの有無をそれぞれ示す。RPやSPが「必須」、「推奨」、「受諾」の場合は「あり」、いずれも掲載されていない場合は「なし」とし

4 章 データ公開の外的要因と日本の研究者の状況

た。以下では、「あり」の比率を「掲載率」と記す。

RP の掲載率は学協会が 52.9%，商業出版社が 67.7%であり，商業出版社の方が高かった。SP は学協会が 86.0%，商業出版社が 93.9%であり，同じく商業出版社の方が高かった。

表 4-18 学協会・商業出版社別データ公開ポリシーの掲載率

出版者	雑誌数	リポジトリ (RP)		補足資料 (SP)	
		あり	なし	あり	なし
学協会	121	64	57	104	17
		52.9%	47.1%	86.0%	14.0%
商業出版社	99	67	32	93	6
		67.7%	32.3%	93.9%	6.1%
合計	220	131	89	197	23
		59.5%	40.5%	89.5%	10.5%

表 4-19 に，学協会と商業出版社別のポリシーの強度を示す。RP が「必須」の比率は学協会が 38.8%，商業出版社が 53.5%であり，「必須」と「推奨」の合計は，学協会が 52.8%，商業出版社が 66.6%であった。いずれも商業出版社の比率が高かった。

SP が「必須」の比率は学協会が 9.1%，商業出版社が 20.2%であり，「必須」と「推奨」の合計は，学協会が 25.6%，商業出版社が 29.3%であった。いずれも商業出版社の比率がやや高かった。

表 4-19 学協会・商業出版社別データ公開ポリシー

出版者	雑誌数	リポジトリ (RP)				補足資料 (SP)			
		必須	推奨	受諾	なし	必須	推奨	受諾	なし
学協会	121	47	17	0	57	11	20	73	17
		38.8%	14.0%	0.0%	47.1%	9.1%	16.5%	60.3%	14.0%
商業出版社	99	53	13	1	32	20	9	64	6
		53.5%	13.1%	1.0%	32.3%	20.2%	9.1%	64.6%	6.1%
合計	220	100	30	1	89	31	29	137	23
		45.5%	13.6%	0.5%	40.5%	14.1%	13.2%	62.3%	10.5%

表 4-20 に分野グループ別の結果を示す。生物・医学は 100 誌のうち学協会の雑誌が 46 誌 (46.0%) であり，自然科学は 90 誌のうち学協会の雑誌が 55 誌 (61.1%) であった。RP，SP によるポリシーの掲載率は，生物・医学，自然科学ともに商業出版社の方が高かった。

表 4-20 学協会・商業出版社別データ公開ポリシー（分野グループ別）

分野 グループ	出版者	雑誌 数	リポジトリ (RP)				補足資料 (SP)			
			必須	推奨	受諾	なし	必須	推奨	受諾	なし
生物・医学	学協会	46	27	6	0	13	5	3	34	4
			58.7%	13.0%	0.0%	28.3%	10.9%	6.5%	73.9%	8.7%
	商業 出版社	54	37	7	1	9	11	6	36	1
			68.5%	13.0%	1.9%	16.7%	20.4%	11.1%	66.7%	1.9%
自然科学	学協会	55	15	10	0	30	4	15	29	7
			27.3%	18.2%	0.0%	54.5%	7.3%	27.3%	52.7%	12.7%
	商業 出版社	35	12	5	0	18	7	2	23	3
			34.3%	14.3%	0.0%	51.4%	20.0%	5.7%	65.7%	8.6%
合計		190	91	28	1	70	27	26	122	15
			47.9%	14.7%	0.5%	36.8%	14.2%	13.7%	64.2%	7.9%

生物・医学について、RP の掲載率（全体から「(ポリシー) なし」を引いた値）は、学協会の雑誌は 71.7%，商業出版社の雑誌は 83.3%であった。SP の掲載率は、学協会の雑誌は 91.3%，商業出版社の雑誌は 98.1%であった。RP の強度については、「必須」の比率は学協会の雑誌が 58.7%，商業出版社の雑誌は 68.5%であり、「必須」と「推奨」の合計は学協会の雑誌は 71.7%，商業出版社の雑誌は 81.5%であった。SP の強度については、「必須」の比率は学協会の雑誌が 10.9%，商業出版社の雑誌は 20.4%であり、「必須」と「推奨」の合計は学協会の雑誌は 17.4%，商業出版社の雑誌は 31.5%であった。

自然科学について、RP の掲載率は、学協会の雑誌は 45.5%，商業出版社の雑誌は 48.6%であった。SP の掲載率は、学協会の雑誌は 87.3%，商業出版社の雑誌は 91.4%であった。RP の強度は「必須」の比率は学協会の雑誌が 27.3%，商業出版社の雑誌は 34.3%であり、「必須」と「推奨」の合計は学協会の雑誌は 45.5%，商業出版社の雑誌は 48.6%であった。SP の強度は「必須」の比率は学協会の雑誌が 7.3%，商業出版社の雑誌が 20.0%であったが、「必須」と「推奨」の合計は学協会の雑誌がやや高く 34.6%，商業出版社の雑誌は 25.7%であった。

[3] OA ステータス

表 4-21 に、OA ステータスごとのデータ公開ポリシーの有無の集計結果を示す。RP か SP のいずれか、あるいは両方が掲載されている場合は「あり」、RP も SP も掲載されていない場合は「なし」にカウントした。調査対象である 220 誌のうち、「OA」誌は 15 誌 (6.8%)，APC やエンバーゴ期間といった制約つきで論文が OA となる「制約つき OA」誌

は 125 誌 (56.8%), OA を認めず購読料を必要とする「非 OA」誌は 80 誌 (36.4%) であった。OA 誌は全て (100.0%) ポリシーを掲載していた。また, 制約つき OA 誌は 93.6%が, 非 OA 誌は 90.0%がポリシーを掲載していた。

表 4-21 OA ステータスとデータ公開ポリシーの有無

データ OA	雑誌数	データ公開ポリシー	
		あり	なし
OA	15	15 100.0%	0 0.0%
制約つき OA	125	117 93.6%	8 6.4%
非 OA	80	72 90.0%	8 10.0%
合計	220	204 92.7%	16 7.3%

表 4-22 に, OA ステータスごとの RP と SP の有無をそれぞれ示す。RP や SP が「必須」, 「推奨」, 「受諾」の場合は「あり」, いずれも掲載されていない場合は「なし」とした。以下では, 「あり」の比率を「掲載率」と記す。

RP は OA 誌 (66.7%), 制約つき OA 誌 (61.6%), 非 OA 誌 (55.0%) の順に掲載率が高かった。一方, SP は制約つき OA 誌 (92.8%), 非 OA 誌 (87.5%), OA 誌 (73.3%) の順に掲載率が高かった。

表 4-22 OA ステータスとデータ公開ポリシーの掲載率

OA ステータス	雑誌数	リポジトリ (RP)		補足資料 (SP)	
		あり	なし	あり	なし
OA	15	10 66.7%	5 33.3%	11 73.3%	4 26.7%
制約つき OA	125	77 61.6%	48 38.4%	116 92.8%	9 7.2%
非 OA	80	44 55.0%	36 45.0%	70 87.5%	10 12.5%
合計	220	131 59.5%	89 40.5%	197 89.5%	23 10.5%

表 4-23 に、OA ステータスごとのポリシーの強度を示す。RP が「必須」の比率は OA 誌 (60.0%)、非 OA 誌 (46.3%)、制約つき OA 誌 (43.2%) の順に高かった。RP が「推奨」または「推奨」の比率は OA 誌 (66.7%)、制約つき OA 誌 (61.6%)、非 OA 誌 (53.8%) の順に高かった。SP は「必須」の比率は非 OA 誌 (23.8%)、制約つき OA 誌 (8.8%)、OA 誌 (6.7%) の順に高かった。「必須」または「推奨」の比率は OA 誌 (40.0%)、非 OA 誌 (36.3%)、制約つき OA 誌 (20.0%) の順に高かった。

表 4-23 OA ステータスとデータ公開ポリシーの強度

OA ステータス	雑誌数	リポジトリ (RP)				補足資料 (SP)			
		必須	推奨	受諾	なし	必須	推奨	受諾	なし
OA	15	9	1	0	5	1	5	5	4
		60.0%	6.7%	0.0%	33.3%	6.7%	33.3%	33.3%	26.7%
制約つき OA	125	54	23	0	48	11	14	91	9
		43.2%	18.4%	0.0%	38.4%	8.8%	11.2%	72.8%	7.2%
非 OA	80	37	6	1	36	19	10	41	10
		46.3%	7.5%	1.3%	45.0%	23.8%	12.5%	51.3%	12.5%
合計	220	100	30	1	89	31	29	137	23
		45.5%	13.6%	0.5%	40.5%	14.1%	13.2%	62.3%	10.5%

4.2.3 分野別ポリシーとデータ公開経験の比較

本章で調査した雑誌のポリシーによるデータ公開要求に対して、日本の研究者はどの程度データを公開しているのだろうか。分野別のポリシーの掲載率と前章で調査したデータの公開経験がある回答者の比率（以下、「公開率」）に関連があるかどうかを比較した。

分野別のポリシーの掲載率(x)と、データ公開率(y_1)の相関分析を行ったところ、正の相関がみられた (Pearson の相関係数 $r=0.525$, $p<0.05$)。つまり雑誌のポリシーの掲載率が高い分野ほどデータ公開率が高い傾向が確認された。図 4-4 に、散布図を示す。また、 x が「分野別のデータ公開要求の状況」を示す説明変数として妥当であるかどうかを確認するために、データ公開経験の理由として雑誌のポリシーを選択した回答者の比率(y_2)との相関分析を行ったところ、同じく正の相関がみられた (Pearson の相関係数 $r=0.624$, $p<0.01$)。

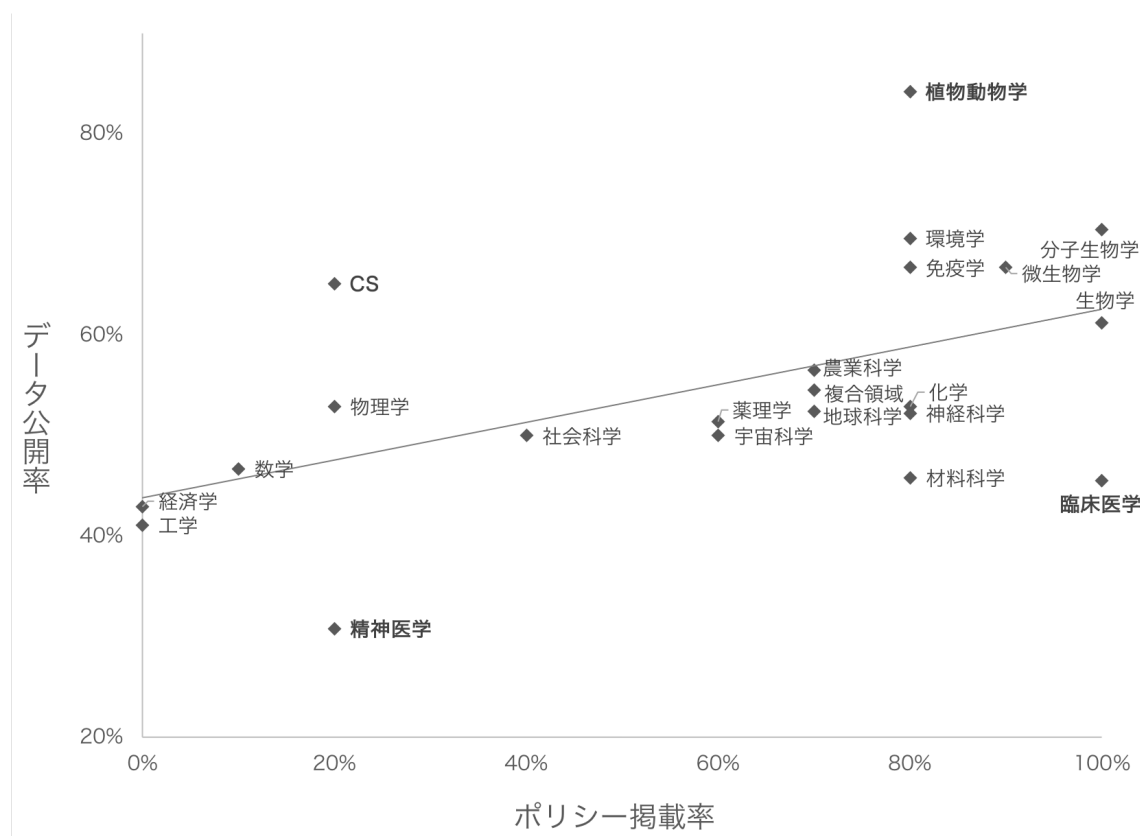


図 4-4 分野別雑誌のデータ公開ポリシーの掲載率とデータ公開率の関連

以上の結果から、分野別のポリシーの掲載率と日本の研究者によるデータ公開率には一定の関連があると判断した上で、ポリシーの掲載率に対して公開率が高い分野と低い分野を特定した。ポリシーの掲載率と比較してデータ公開率が高い分野，すなわち，回帰式 $y=0.1878x+0.4380$ の外れ値⁴⁰⁾を確認すると，植物・動物学とコンピュータサイエンスであった（図 4-4）。逆にポリシーの掲載率と比較してデータ公開率が低い分野は，臨床医学と精神医学・心理学であった。たとえば臨床医学で IF が高い原著論文誌 10 誌は，全てデータ公開ポリシーを掲載していた（表 4-5）のに対して，データ公開率は 45.5%にとどまっていた（図 3-6）。

4.3 考察

ここでは，(1)分野別のポリシーの掲載率と要求の強度，(2)雑誌の特徴とポリシーの掲載率と要求の強度の関連，(3)雑誌のポリシーと日本の研究者のデータ公開状況，(4)雑誌のポリシ

⁴⁰⁾ 4.1.6 で述べたように，本研究では残差が 15%以上または-15%以下の場合を外れ値とした。

一調査から得られた知見に基づく課題について、それぞれ考察する。

4.3.1 ポリシーの掲載率と要求の強度

雑誌によるデータ公開ポリシーを、22 分野の IF 上位各 10 誌（合計 220 誌）を対象として調査した結果、RP または SP を掲載していた雑誌は 204 誌（92.7%）であった。表 2-9 に示した先行研究と比較して、最も掲載率が高かった。これは本研究よりも IF が高い雑誌を対象とした Alsheikh-Ali et al. (2011) の 88.0%をも上回る結果であった。計算科学を対象とした Stodden et al. (2013) によれば、2011 年（32.9%）よりも 2012 年（37.6%）の方がポリシーの掲載率が高かった。また、生物学分野を中心とした調査を比較すると、2007 年は 60.0%（Piwowar & Chapman, 2008）、2009 年は 88.0%（Alsheikh-Ali et al., 2011）であった。以上のことから、調査年が新しいほどポリシーの掲載率が高い可能性が示唆された。

RP は 220 誌のうち 131 誌（59.5%）が掲載しており、130 誌（59.1%）は要求の強度が高い「必須」または「推奨」であった。一方、SP は 197 誌（89.5%）が掲載しており、より多くの雑誌が採用していたが、137 誌（62.3%）は要求の強度が低い「受諾」であった。RP と SP は掲載率も強度の分布も異なることが明らかになった。

RP の掲載率と強度は分野によるばらつきがみられた（図 4-2）。先行研究から、分野ごとの IF や雑誌年齢の分布がポリシーの掲載率や要求の強度に影響を与えている可能性があると考えられるため、相関分析を行って関連を調べた。その結果、IF は、やや影響がみられた（掲載率や要求の強度が高い分野は、比較的 IF が高い傾向にある）が、強い相関はみられなかった。また、雑誌年齢はほとんど相関がみられなかったため、雑誌を再選定せずに分野間比較を行うこととした。

RP の掲載誌数が 10～8 誌であった「グループ A」の 10 分野は、調査対象誌（各 10 誌）のうちのいずれかが、表 4-10 に挙げたりポジトリをデータの公開先として指定していた。表 4-10 に挙げたりポジトリは、Dryad を除き 10 年以上の歴史を持つ世界規模のデータリポジトリであり、多くの機関による認証・評価基準を満たしている（Day, 2008; Center for Research Libraries, n.d.）。Dryad は開始年こそ新しいものの、大規模かつ永続性のあるリポジトリとして米国科学財団（National Science Foundation, NSF）等から助成を受けて運営されている。2014 年度には 200 誌以上の雑誌に掲載された論文の 2,714 件のデータパッケージが登録された（Dryad, 2014）。つまり、この 10 分野はデータの公開先として、学術コミュニティに認知された「定評のあるリポジトリ」が存在することが明らかになった。

表 4-24 に、オンラインのデータリポジトリ目録 Databib⁴¹⁾による分野別データリポジトリ数を示す。ここでは、考察の参考とするため調査時点に近い 2014 年 12 月の値を用いた。

⁴¹⁾ 2015 年に Databib は re3data.org とともに DataCite に統合された。

4 章 データ公開の外的要因と日本の研究者の状況

Databib に収録されたデータリポジトリは、複合領域を除くと生物科学（17.5%）、地球科学（11.8%）、数学・物理学（10.7%）の順に多く、工学や経営学は少ないことがわかる。

表 4-24 分野別データリポジトリ数（Databib）

分野	リポジトリ数
複合領域	179 (18.0%)
生物科学	174 (17.5%)
地球科学	117 (11.8%)
数学・物理学	106 (10.7%)
社会科学	92 (9.3%)
環境科学	87 (8.8%)
健康・医学	60 (6.0%)
農業	18 (1.8%)
生態系科学	15 (1.5%)
言語学・文学	11 (1.1%)
その他（工学，経営学ほか）	48 (4.8%)
未分類	85 (8.6%)
合計	992 (100%)

調査日：2014 年 12 月 24 日

予備調査と同様に、生物学や医学の関連領域は RP の掲載率も強度もおおむね高い傾向にあり、定評のある生物データのリポジトリへの公開を求めている。表 4-24 から、生物科学、健康・医学分野のリポジトリ数が比較的多いこともわかる。ただし、農業科学は 7 誌、薬理学・毒性学は 6 誌、精神医学・心理学は 2 誌しか掲載しておらず、領域内で差がみられた。これら 3 分野のポリシーの掲載率が低い理由として、企業の利益保護のためにデータが非公開とされてきたことや、被験者や患者のプライバシー保護などの問題が考えられる。しかし、3 章の質問紙調査の結果を確認すると、「商用利用される可能性」への懸念が強い順に、薬理学・毒性学が 5 位（「問題である」と「やや問題である」の合計は 77.2%）、精神医学・心理学が 11 位（同 69.3%）、農業科学が 12 位（同 67.4%）であった（図 3-10）。平均は 69.8%であり、この 3 分野が特に強いとはいえない。また、「機密・プライバシー情報」への懸念は強い順に、薬理学・毒性学が 5 位（「問題である」と「やや問題である」の合計は 80.0%）、精神医学・心理学が 8 位（同 76.9%）、農業科学が 16 位（同 60.9%）であった（図 3-11）。平均は 71.8%であり、同じく強いとまではいえない。つまり、他の生物・医学分野でも同じように問題になっていると考えられる。一方、これらの分野の懸念や不足感が強かったのは「データ公開用リポジトリの充足度」であり、精神医学・心理学が 1 位（「不十分」と「やや不充

分」の合計は 92.3%)、農業科学が 4 位 (同 63.0%)、薬理学・毒性学が 5 位 (同 62.8%) であった (図 3-8)。全体平均は 54.6%であり、相対的に充足度が低いと認識されていた。表 4-24 の Databib の収録リポジトリは本研究とは分野分類が異なるが、「農業」のリポジトリが比較的少ない。以上のことから、生物・医学分野で RP の掲載率が低い分野は、公開用リポジトリが不十分である可能性が示唆された。

RP の掲載誌数が 7~4 誌であった「グループ B」について、公開データ (一次データ) の二次分析を行ってきた (Borgman, 2015) 地球科学、宇宙科学、社会科学は一次データへのリンクを求める記述もみられた (4.2.1)。しかし、RP の掲載率や強度はそれほど高くなく、二次分析したデータの公開を求める雑誌は一部に留まっていた。図 4-4 に示した RP の強度に対するデータ公開率もほぼ同程度であった。

RP の掲載誌数が 2~0 誌であった「グループ C」について、物理学と数学は、Databib の収録リポジトリ数が全 992 件のうち 106 件であり、比較的多かった (表 4-24)。しかし RP の掲載誌数と強度は、物理学が「必須」2 誌、数学が「推奨」1 誌のみであり、いずれも低かった。しかし、物理学の雑誌の投稿規定は、RP が「なし」の場合であってもデータに関する記述がみられた。また、物理学の対象誌は実験や観察によるデータ作成や収集を伴う研究を含んでいた。一方、数学の対象誌はデータを用いない理論研究が多かったため、やや強度が低かったと考えられる。質問紙調査の自由回答でも、“数学は純粋に理論的な研究分野であるため、実証研究的な意味合いでのデータというものとは存在せず、(数表・分類表などのように) 特定の数学的事実を網羅的にまとめたものや、抽象的・一般的事実の具体例を提示するものとしてのみデータが扱われることになる”という指摘がみられた (3.2.8)。

同じく「グループ C」のコンピュータサイエンス、工学、経済学・経営学は Databib の収録リポジトリ数が少なく、コンピュータサイエンス (Databib の分類では「コミュニケーションと情報学」) が 3 件、工学が 8 件、経営学・経営学 (同「経営学」) が 2 件であった。これら 3 分野の投稿規定はデータに関する記述を含んでおり、研究内容によってデータが用いられていると推測される。しかし、RP の掲載誌数を確認すると、コンピュータサイエンスは 2 誌、工学と経済学・経営学は 0 誌であった。つまり、この 3 分野は数学とは異なり、データを用いた研究が行われているにもかかわらず、RP の掲載率が低かったといえる。

4.3.2 雑誌の特徴とポリシーの関連

先行研究と予備調査の結果を参考に、雑誌の特徴として倫理と COI の開示に関する記述、出版者、OA ステータスを調査して、ポリシーの掲載率や強度に違いがあるかどうかを検証した。ここでは、それぞれの理由を考察する。

[1] 倫理と COI の開示

「研究倫理や COI の開示に関する記述がある雑誌はポリシーの掲載率や要求の強度が高い」という仮説は支持された。表 4-13 に示したように、倫理に関する記述有りの雑誌は 97.7% が、無しの雑誌は 85.7% がデータ公開ポリシーを掲載していた。また、COI の開示に関する記述有りの雑誌は 97.2%、無しの雑誌は 84.4% が掲載しており、いずれも記述がある雑誌の方がポリシーの掲載率が高かった。

同様に、倫理や COI の記述ありの雑誌は RP、SP の掲載率も高く（表 4-14）、強度についても「必須」とする雑誌の割合、および「必須」または「推奨」とする雑誌の割合が高かった（表 4-15）。分野グループ別にみると、倫理も COI に関する記述も生物・医学の方が自然科学分野よりも掲載率が高かった。表 4-16 で RP と SP の掲載率、および RP と SP の強度を確認すると、分野グループ別にみても倫理や COI の開示の記述有りの雑誌の方が、掲載率も強度も高い傾向にあった。唯一、生物・医学グループは RP を「必須」とする雑誌の比率が倫理の記述無し（66.7%）の方が有りよりも高かったが、「必須」と「推奨」の合計は、有り（80.6%）の方が無し（69.7%）よりも高かった。従って、分野によるばらつきを排しても、倫理や COI の開示に関する記述有りの雑誌の方が、無しの雑誌よりもポリシーの掲載率や要求の強度が高いといえるだろう。

かねてより科学の深刻な問題である（Price, 1964）研究不正への対策は、データ公開の目的の 1 つとして論じられることが多いが、本調査によって倫理や COI に関する記述がポリシーの掲載率や強度と関連があることが示された。直接的な関連を示す例として、*Neurology* は産業界がスポンサーの研究はデータを全て公開するよう求めていた⁴²⁾。また、学術雑誌にとってデータ公開要求が研究不正を低減し、透明性を高めるための取り組みであることを示す例として、2014 年に Transparency and Openness Promotion (TOP) ガイドラインが公開された（Nosek, et al., 2015）。TOP ガイドラインは、研究の透明性を高めるために、学術雑誌によるデータ公開やデータ引用の要求レベルを Level 0（Encourages（推奨））、Level 1（Disclose（公開））、Level 2（Require（要求））、Level 3（Verify（義務化と査読））に分類している。

[2] 出版者

「学協会の雑誌の方が商業出版社の雑誌よりもポリシーの掲載率や要求の強度が高い」という仮説は棄却された。商業出版社の雑誌は 96.0% が、学協会は 90.1% がデータ公開ポリシーを掲載していたため、商業出版社の雑誌の方がデータ公開ポリシーの掲載率が高かった（表 4-17）。RP、SP それぞれの掲載率も同様であり（表 4-18）、さらに「必須」とする雑誌の割合

⁴²⁾ “Full Data Access”. *Neurology*. <http://www.neurology.org/site/misc/auth2.xhtml#FULldataACCESS>, (accessed 2014-5-8).

も高かった（表 4-19）。分野グループ別にみても、商業出版社の雑誌の方が掲載率も強度も高い傾向にあった。ただし、自然科学の SP は、「必須」とする雑誌の割合が商業出版社（20.0%）の方が学協会（7.3%）よりも高いものの、「必須」と「推奨」の合計だけは学協会（34.6%）の方が商業出版社（25.7%）よりも高かった。逆転が起きていたのは「必須」と「推奨」の合計のみであることから、分野によるばらつきを排しても、商業出版社の雑誌の方が学協会の雑誌よりもポリシーの掲載率や要求の強度が高いといえるだろう。出版者にとって補足資料の公開は負担となるため（4.2.1）、商業出版社は SP による要求を控えめにしているのかもしれない。

Piowar and Chapman（2008）と Stodden et al.（2013）の調査では、学協会の方が掲載率が高く、本調査とは逆の結果であった。その原因として、Stodden et al.（2013）は Springer 社や Wiley 社などの商業出版社と学協会の雑誌をそれぞれ集計しているが、商業出版社に学協会の出版物を含めた可能性があると考えられる。本調査の対象誌には商業出版社のプラットフォームを利用した学協会の雑誌が 46 誌（全体の 20.9%）含まれていた⁴³⁾ため、これらは学協会として集計する方が妥当であると判断した。

本調査では、商業出版社の雑誌の方が学協会の雑誌よりもポリシーの掲載率や要求の強度が高かった要因までは明らかにできなかったが、商業出版社は雑誌の信用度やブランドを維持するために研究不正対策としてデータ共有を強く要求しているという可能性が考えられる。証左の一例として、有力誌ほど論文の撤回が多く、また撤回の通知は不透明な場合が多いが、その理由は著者や雑誌が体面を保つ、あるいは名誉毀損訴訟を避けるためではないかと推測されている（Corbyn, 2012）。また、本研究は IF が高い雑誌を対象としたが、IF の高さ論文の撤回率には強い正の相関があるとする研究（Fang, Steen, & Casadevall, 2012）や、IF が高い雑誌ほど研究不正があった論文の撤回までの時間が短いとする調査（Steen, Casadevall, & Fang, 2013）がある。つまり IF が高い雑誌ほど、研究不正対策を強く意識しており、データ公開ポリシーの強度が高いのではないかと考えられる。IF が低～中程度の雑誌のポリシーを調査して比較することを今後の課題としたい。

[3] OA ステータス

「OA 誌はポリシーの掲載率や要求の強度が高い」という仮説は部分的に支持された。OA 誌は論文と同様にデータも公開するよう求めているのかどうかを確認するために、OA ステータスとポリシーの掲載率や強度を比較した。その結果、OA の 15 誌は全てポリシーを掲載しており（100.0%）、制約つき OA 誌（93.6%）や非 OA 誌（90.0%）と比較して掲載率が高かった（表 4-21）。ただし、全体的にポリシーの掲載率が高かったこと（92.7%）、OA 誌の比率

⁴³⁾ Wiley 社は 31 誌中 16 誌、Elsevier 社は 26 誌中 7 誌、Springer 社は 16 誌中 10 誌、Oxford University Press は 11 誌中 8 誌、Sage Publications 社は 4 誌中 3 誌が学協会による雑誌であった。

が低かったこと（全体の 6.8%）から、本研究の結果によって OA 誌の方が掲載率が高いとまでは言い切れなかった。OA ステータスとポリシーの掲載率の関連を明らかにするためには、調査対象に OA 誌を増やして差があるかどうかを検証する必要があると考えられる。

OA 誌は、RP の掲載率（66.7%）が制約つき OA 誌（61.6%）や非 OA 誌（55.0%）と比較して高かった（表 4-22）。また、強度についても「必須」の比率（60.0%）が制約つき OA 誌（43.2%）や非 OA 誌（46.3%）と比較して高く、「必須」と「推奨」の合計も同様であった（表 4-23）。以上のことから RP については、OA 誌の掲載率や要求の強度が高い可能性が示唆された。

SP の掲載率（73.3%）は、制約つき OA 誌（92.8%）や非 OA 誌（82.5%）の方が高く、「必須」の比率は非 OA 誌（23.8%）、制約つき OA 誌（8.8%）、OA 誌（6.7%）の順に高かった。データ公開の背景として OA の理念が述べられることが多いが、OA ステータスは SP の掲載率や強度と関連があるとはいえなかった。

Piowar and Chapman（2008）の調査では、OA 誌かどうかポリシーの強度に影響を与えていることが示されていた。本研究では RP については同様であったものの、SP については結果が異なった。本調査ではその理由までは明らかにできないが、非 OA 誌は論文を公開しないかわりに補足資料としてデータを公開することによって研究の透明性を担保しようとしているのではないだろうか。また、補足資料を公開する場合の負担感（4.2.1）は OA 誌の方が大きいのではないかと推測される。これらの検証は今後の課題としたい。

4.3.3 雑誌のポリシーの課題

ポリシーの調査過程で得られた知見をふまえて、よりスムーズにデータを公開できるようなポリシーのあり方について考察する。まず定形ポリシーの策定について、続いてデータを公開するリポジトリについて述べる。

[1] 定形ポリシーの策定

本研究で調査したデータ公開ポリシーの多くは雑誌が独自に作成しており、投稿規定の中の記載箇所も記載内容もそれぞれ異なっていた。一般に、研究者は研究内容に応じてさまざまな雑誌に論文を投稿すると考えられるが、データ公開の対応のみならず、その理解にも時間や手間などのコストがかかっていると推測される。一方、Elsevier 社や Nature Publishing Group (NPG)⁴⁴⁾などの出版社には定形のポリシーがあり、それぞれ複数の雑誌が採用していた。共通の定形ポリシーがあれば、研究者は投稿先によってポリシーを理解したりデータ公開の対応を変えたりする必要がなくなり、雑誌の発行者はポリシー策定のためのコストが削

⁴⁴⁾ 2015 年から Springer Nature 社。

減できるため、双方にとって有益であると考えられる。

倫理や COI の開示については、4.2.2 に示したように既存の COPE ガイドライン、CONSORT 声明、ヘルシンキ宣言、ICMJE 勧告などを採用している雑誌が多くみられた。データ公開ポリシーの雛形を示した例として、2011 年に Dryad は進化学分野を対象とした Joint Data Archiving Policy (JDAP) を公開した (Dryad, 2011)。具体的には、“[雑誌名] は、出版の条件として、論文の結果を補助するデータは [承認したアーカイブのリスト] といったパブリックアーカイブに登録することを求める。...”といった記述例が示されている。また、Joint Information Systems Committee (JISC) ⁴⁵⁾ の助成を受けた Journal Research Data (JoRD) プロジェクトによって学術雑誌のデータ公開ポリシーのモデル (Sturges et al., 2015; Moles, 2015) や 8 項目の原則 (Starr et al., 2015) が示されている。2014 年に公開された TOP ガイドライン (4.3.2) も、データ公開やデータ引用ポリシーの雛形として使用することができる (Center for Open Science [COS], 2015)。2017 年 4 月には研究データ同盟 (Research Data Alliance, RDA) に Data policy standardisation and implementation の Interest Group (IG) が設置され、学術雑誌を中心に、助成機関などでも活用できる研究データポリシーの標準化に向けた取り組みを開始している⁴⁶⁾。3 章の質問紙調査によれば、日本の研究者はデータを公開する場合に時間が不足しているという認識が強かった (図 3-1)。標準化された定形ポリシーが出版者や助成機関、研究機関に広く普及すれば、研究者は論文の投稿先のみならず、助成金の申請先によって、あるいは所属機関を移動しても対応を変える必要がなくなるため、効率的にデータ公開に取り組むことができると考えられる。

[2] データの公開先リポジトリ

データを公開するリポジトリについて、RP を掲載する 131 誌のうち 7 誌はデータの公開先を指示していなかった (4.2.1)。また、OA 誌を対象としたデータ公開ポリシーの調査によれば、50 誌のうちデータの登録場所を特定していたのはわずか 8%であった (Castro, Crosas, Garnett, Sheridan, & Altman, 2017)。しかし、研究者にとって時間がないことや公開用のリポジトリがないことは、データ公開の障壁となっている (表 2-8, 図 3-1, 図 3-2)。Retrovirology のように既存のディレクトリからリポジトリを選ぶよう指示すれば、研究者がリポジトリ選択にかかる時間や公開を諦めるケースを低減させることができるのではないだろうか。

補足資料の掲載にはコストがかかるため、多くの雑誌はデータ量やファイル数を制限しており、課金を行う場合もみられた (4.2.1)。一方、Knowledge Exchange は購読型の雑誌や論文であっても補足資料を無料で公開することを出版社に推奨している (van den Eynden & Bishop,

⁴⁵⁾ 現在は Jisc。

⁴⁶⁾ Data policy standardisation and implementation IG. <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation>, (accessed 2018-10-31). RDA については、村山 (2016) に詳しい。

2014)。補足資料の問題への対策として、*Systematic Biology* のように永続性の高い (Anderson et al., 2006) 外部リポジトリに補足資料を登録する方法は、研究者と出版者の双方にとって便益が高いと考えられる。こうしたリポジトリに登録することは、補足資料の問題点として指摘されているリンク切れ (van den Eynden & Bishop, 2014; Evangelou et al., 2005; Anderson et al., 2006; Pepe et al., 2014) などの問題への対応にもなるだろう。また、Cell Press のように補足資料は必要最低限の量に留めることも重要であると考えられる。データ公開ポリシーに、公開すべきデータに関するガイドラインを盛り込むことによって投稿者の判断が容易になると考えられる。雑誌のポリシーではないが、参考例として、英国の Digital Curation Centre (DCC) は、“*Five steps to decide what data to keep: A checklist for appraising research data* (保存すべきデータを選択するためのガイド)”を公開している (Whyte, 2014)。

4.3.4 雑誌のポリシーと日本の研究者のデータ公開状況

雑誌のポリシーによる要求 (RP の掲載率) と、前章で明らかにした日本の研究者によるデータ公開率を分野別に比較した。両者には正の相関がみられ、Vines et al. (2013b) を支持する結果であった。RP の掲載率から期待されるデータ公開率よりも実際の公開率の方が高かったのは植物・動物学とコンピュータサイエンスであり、公開率が低かったのは臨床医学と精神医学・心理学であった。植物・動物学とコンピュータサイエンスは雑誌の要求によらず、自発的にデータ公開を行っていると考えられ、臨床医学は雑誌の要求があるにもかかわらずデータ公開が困難であると考えられる。そして精神医学・心理学は、雑誌による要求の強度もデータ公開率も他の分野と比較して低かった。ここでは、植物・動物学とコンピュータサイエンスを「自発的にデータ公開を行っている分野」、臨床医学と精神医学・心理学を「データ公開が困難な分野」とする。これらの分野は、どのようなデータ公開の動機や障壁、特徴があるのだろうか。

表 4-25 に、3 章の結果のうち分野による差がみられた項目について、4 分野の結果をまとめた。データの提供・被提供の頻度には、「よくある」と「たまにある」の選択率の合計を示した。カレントデータを公開する場合の障壁の程度は、「不十分」と「やや不十分」、および「問題である」と「やや問題である」の選択率の合計を示した。

表 4-25 分野による差がみられた項目と特徴的な分野の結果

質問項目	平均	植物	CS	臨床	精神	
データの公開状況						
Q18 データの公開経験	51.2%	84.2%	65.1%	45.5%	30.8%	
Q2 OA の経験	70.9%	84.2%	74.6%	72.7%	100%	
データの公開理由						
Q19 科学研究や成果実装の推進	26.1%	43.8%	46.3%	20.0%	0.0%	
Q19 他の研究者からのリクエスト	25.4%	37.5%	34.1%	30.0%	25.0%	
Q19 オープンデータへの貢献	10.5%	37.5%	22.0%	0.0%	0.0%	
データの非公開理由と意思						
Q20 時間が必要だから	25.9%	33.3%	40.9%	33.3%	44.4%	
Q20 ニーズがないと思うから	22.1%	33.3%	9.1%	16.7%	22.2%	
Q20 リポジトリなどの公開手段がない	8.9%	0.0%	18.2%	8.3%	44.4%	
Q21 データの公開意思	28.7%	33.3%	35.3%	33.3%	33.3%	
データの提供・被提供状況						
Q9 データの提供頻度	44.2%	57.9%	52.4%	36.4%	23.1%	
Q10 データの被提供頻度	38.9%	63.2%	55.5%	27.3%	15.4%	
公開データの入手経験						
Q11 公開データの入手経験	76.0%	89.5%	85.7%	50.0%	46.2%	
Q13 再分析・再利用を行う	16.6%	47.1%	29.6%	0.0%	0.0%	
カレントデータのプロファイル						
Q25 データの機密情報	49.3%	5.3%	47.5%	68.2%	69.2%	
カレントデータを公開する場合の障壁の程度						
Q28 公開の資源	研究中のストレージ	47.8%	42.1%	36.5%	54.5%	76.9%
	公開用リポジトリ	54.6%	42.2%	50.8%	59.1%	92.3%
Q29 データ公開の懸念	商用利用	69.8%	31.6%	58.7%	81.8%	69.3%
	機密・プライバシー	71.8%	42.1%	63.4%	90.9%	76.9%
	研究の誤りを発見	14.7%	5.3%	11.1%	18.2%	15.4%
	引用せずに利用	87.7%	89.5%	82.6%	77.2%	92.3%
	先に論文を出版	84.7%	94.8%	74.6%	77.2%	84.6%

自発的にデータ公開を行っている分野について、植物・動物学のデータの公開理由は、表 4-25 に示した全体平均と比較すると、「オープンデータへの貢献」や「科学研究や成果実装の推進」の選択率が高かった。また、データの提供頻度、被提供頻度、データの再分析・再利

用を行う頻度が高かった。一方、データに機密情報を含む回答者の比率、商用利用への懸念、機密・プライバシー情報への懸念が低かった。

コンピュータサイエンスのデータの公開理由では、「科学研究や成果実装の推進」の選択率が高かった。また、データの提供頻度、被提供頻度も高かった。表 4-25 に挙げたデータ公開の障壁は、全て平均よりも低かった。

以上のことから、自発的にデータ公開を行っている植物・動物学とコンピュータサイエンスは、機密や商用利用などへの懸念が低く、データのやりとりを行っていることがわかった。データ公開の理由として「科学研究や成果実装の推進」が多く選ばれるなど、科学的な利他性も醸成されていると考えられる。また、表 3-25 に示したデータ公開の理由のうち、全体5位の「所属機関」(23.9%)は、コンピュータサイエンスは 17.1% (13 位)、植物・動物学は 6.3% (20 位)であり、制度はこれらの分野の主要な公開理由とはいえないことがわかる。このことから、植物・動物学とコンピュータサイエンスは、自発的にデータを公開している傾向があるといえよう。

データ公開が困難な分野について、臨床医学はデータの提供頻度や公開データの入手経験がある回答者の比率が低く、再分析・再利用を行う回答者はいなかった。データに機密情報を含む回答者の比率は高く、機密・プライバシーへの懸念も高かった。一方、2017 年 6 月に ICMJE は「臨床試験データ共有に関する声明：医学雑誌編集者国際委員会の要件 (Data sharing statements for clinical trials: A requirement of the International Committee of Medical Journal Editors)」を公開した (Taichman, et al., 2017)。声明には、(1)2018 年 7 月 1 日以降、ICMJE 会員誌に投稿される臨床試験の結果を報告する論文にはデータ共有ステートメントを含まなければならないこと、(2)2019 年 1 月 1 日以降に参加者登録を開始する臨床試験は、データ共有計画を含めなければならないことが明示されている⁴⁷⁾。この声明を紹介した栗原と齊尾 (2017) は、“臨床試験において、データ共有は比較的新しい動きであり、日本ではまだ、研究者たちに十分に周知されているとはいえない” (p. 488) と指摘している。雑誌のみならず、日本医療研究開発機構 (AMED) (2018) によるデータマネジメントプランの義務化などの動きもあることから、早急な課題解決が必要であると考えられる。

精神医学・心理学は、データの公開理由では、「科学研究や成果実装の推進」の選択率が低く、データの非公開理由では、「リポジトリなどの公開手段がない」や「時間が必要だから」の選択率が高かった。データの提供頻度、被提供頻度、入手経験は低く、再分析・再利用を行う回答者はいなかった。従って、他分野と比較して、個人的なデータのやりとりも、公開データの活用も行われていないと推測される。データに機密情報を含む回答者の比率が高く、公開用リポジトリの不足度や研究中のストレージの不足度も高かった。ただし、非公開理由

⁴⁷⁾ “Clinical Trials”. International Committee of Medical Journal Editors (ICMJE).
<http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html>,
(accessed 2018-10-31).

が解決した場合にデータを公開しようとする回答者の比率は 33.3%であり、全体平均の 28.7%よりも高かった。また、心理学分野のデータを使いたいと考える他分野の研究者は 5.3% (73 名) であった (表 4-16)。精神医学・心理学は、データ共有の進展を目指すとする論考もみられる (Wichert, Borsboom, Kats, & Molenaar, 2006; Perrino, 2013; Poldrack, 2015)。以上のことから、臨床医学や精神医学・心理学は、主に機密・プライバシー情報の問題からデータ公開が困難な分野であると考えられるが、公開に向けた動きもみられる。従って、機密を保持しつつデータを公開するための環境が整えば、公開が進む可能性もあると考えられる。

4.3.5 日本におけるデータ公開の課題と展望

ここでは 3 章と 4 章の結果から、日本の研究者によるデータ公開の動機と障壁と「自発的にデータ公開を行っている分野」および「データ公開が困難な分野」の動機や障壁を総合して、日本におけるデータ公開の主要な課題として考察する。続いて、データを公開している日本の研究者の特徴と「自発的にデータ公開を行っている分野」および「データ公開が困難な分野」のデータ公開経験と関連がみられた項目を総合して、データ公開経験や態度と関連がある経験・認識として考察する。最後にまとめとして、データ公開の主要な課題に対する方策と期待される効果について述べる。

[1] 日本におけるデータ公開の主要な課題

3.3.2 で述べたように、データの非公開理由のうち選択率が高い項目や解決が公開に繋がると推測できるような項目はなかった。また、データを公開する場合の資源の充足度はおしなべて低く、データを公開する場合の懸念は「研究の誤りを発見する可能性」を除いていずれも強かった。そこで次の(A)～(E)に 2 つ以上あてはまる項目を、日本におけるデータ公開の主要な課題であるとみなすこととした。すなわち、(A)データの公開理由としてあまり選ばれなかった項目 (選択率 10%未満)、(B)データの非公開理由 (選択率 20%以上⁴⁸⁾)、(C)データ公開の障壁 (充足度が低い資源 3 件と懸念が強い項目 3 件)、(D)データ公開経験がある回答者にとって低い障壁 (充足している資源と懸念が低い項目)、(E)データ公開に消極的な回答者にとって高い障壁 (充足度が低い資源と懸念が強い項目)、(F)自発的にデータ公開を行っている分野の研究者にとって低い障壁、(G)データ公開が困難な分野の研究者にとって高い障壁である。以上の条件から選択された重要な課題は、「データ公開が業績とならないこと」、「時間」、「人材」、「資金」、「引用せずに利用される可能性」、「先に論文を出版される可能性」、「商用利用される可能性」、「機密・プライバシー情報」、「誤解や誤用の可能性」、「公開用リポジト

⁴⁸⁾ 表 3-10 に示したように、非公開理由は 1 位から 5 位までが 20%台であり、6 位が 8.9%であったため 20%を区切りとした。

4 章 データ公開の外的要因と日本の研究者の状況

り」であった。

表 4-26 データ公開の主要な動機と障壁

要因	A	B	C	D	E	E	F
学術雑誌のポリシー		26.4%					
助成機関のポリシー	6.3%						
所属機関のポリシー		22.4%					
分野・コミュニティの規範	9.5%						
業績	9.3%	21.7%					
他の研究者からの要求		22.1%					
科学研究や成果実装の推進							
研究成果の認知度向上							
オープンデータへの貢献							
研究中のストレージ							✓
公開用リポジトリ				✓			✓
保存用ストレージ							
時間		25.9%	74.5%	✓			✓
人材			79.1%	✓			
資金			74.1%	✓			
商用利用される可能性				✓	✓	✓	
機密・プライバシー情報				✓		✓	✓
誤解や誤用の可能性			74.0%		✓		
先に論文を出版される可能性			84.6%		✓		
引用せずに利用される可能性			87.8%		✓		
研究の誤りを発見される可能性							

A＝データの公開理由としてあまり選ばれなかった項目（選択率）、B＝データの非公開理由（選択率）、C＝データ公開の障壁（「不十分」と「やや不十分」、「問題である」と「やや問題である」の合計）、D＝データ公開経験がある回答者にとって低い障壁、E＝データ公開に消極的な回答者にとって高い障壁、F＝自発的にデータ公開を行っている分野の研究者にとって低い障壁、G＝データ公開が困難な分野の研究者にとって高い障壁

以上の結果を日本におけるデータ公開の主要な課題として、学術機関、学術基盤の提供者、学術出版社、助成機関、政府関係者といったステークホルダーが対策を講じる観点から整理

すると、(1)資源に関する課題（公開用リポジトリ，時間，人材，資金），(2)制度に関する課題（業績），(3)技術的解決が可能な課題（機密・プライバシー情報，誤解や誤用の可能性），(4)規範・慣習の醸成が必要な課題（引用，商用利用）に分けられる。以下では，表 4-27 に示した課題を順に考察する。

表 4-27 日本におけるデータ公開の主要な課題

カテゴリ	課題
(1) 資源に関する課題	公開用リポジトリ 公開のための時間，人材，資金
(2) 制度に関する課題	データ公開が業績とならないこと 先に論文を出版される可能性
(3) 技術的解決が可能な課題	機密・プライバシー情報 誤解や誤用の可能性
(4) 規範・慣習の醸成が必要な課題	引用せずに利用される可能性 商用利用される可能性

(1) 資源に関する課題（公開用リポジトリ／時間／人材／資金）

公開用リポジトリの充足度は，「不十分」と「やや不十分」をあわせて 54.6%であった（図 3-1）。また，「わからない」の選択率が 6 項目の中で最も高く（27.1%），認知度が低いこともわかった。分野別にみると，RP と比較してデータ公開率が低かった精神医学・心理学はデータ公開用リポジトリの充足度が最も低く，逆に RP と比較して公開率が高かった植物・動物学は全体の 2 位，コンピュータサイエンスは 3 位であった。

データの公開先（表 3-6）や公開データの入手先（表 3-14）で最も多かったのは，いずれも個人や研究室のウェブサイト，次いで論文の補足資料であった。先行研究（Tenopir et al., 2011; Ferguson, 2014; Kratz & Strasser, 2015a）と同様に，所属機関や特定分野のリポジトリは十分に活用されていないと考えられる。分野別に確認すると，データの公開率が高い分野ほど，公開先や入手先として分野リポジトリを選択していた（表 3-29）。検索ツールとして用いられていたのは，サーチエンジン，次いで論文などの参考文献であった（表 3-17）。検索ツールとして分野リポジトリを選択していたのは，11 分野の回答者のみであった。

以上のことから，リポジトリはデータの公開先としても，入手先としても，検索ツールとしても，充分活用されているとはいえなかった。表 4-28 に，それぞれの選択率と，最も選ばれていたツールを示す。

表 4-28 リポジトリの利用状況

利用方法	機関リポジトリ	分野リポジトリ	最も選ばれていたツール
公開先	34.2%	16.4%	個人や研究室のウェブサイト (50.8%)
入手先	50.0%	23.1%	個人や研究室のウェブサイト (64.8%)
検索ツール	32.7%	2.6%	サーチエンジン (83.6%)

個人や研究室のウェブサイトは、機関や分野のリポジトリと比較すると、永続的に公開することや発見可能性を高めることなどが困難であると考えられる。一方、よく再利用されるデータは、適切に整備されて永続性のあるリポジトリで公開されていることが複数の研究で指摘されている (Horton & Katsanidou, 2011; Faniel, Kriesberg, & Yakel, 2016; Peters, Kraker, Lex, Gumpenberger, & Gorraiz, 2016)。2 章で言及した FAIR 原則 (FORCE11, 2016; 大波ら, 2018) に準じて、Findable (見つけられる), Accessible (アクセスできる), Interoperable (相互運用できる), Re-usable (再利用できる) 状態でデータを公開可能とすることが重要であると考えられる。また、データの入手における問題 (表 3-16) に配慮しつつ、データ公開要求に対してリポジトリの充足度が低い分野に向けたリポジトリ構築を行うことが考えられる。

データを整備し、公開しようとする場合の資源は、特に時間、人材、資金が不足していると認識されていることがわかった。逆にデータを公開している研究者は、公開のための時間の充足度が高かった。1.3.1 に示した通り、時間不足はデータ公開に限らない日本の研究環境の問題となっている。日本の研究者が、データ公開のための時間やデータ公開のための技術や知識を身につけるための時間を新たに確保することは困難であろう。「学術情報のオープン化の推進について (審議まとめ)」では、データ公開について“技術職員、URA 及び大学図書館職員等を中心としたデータ管理体制を構築し、研究者への支援に資するとともに、必要に応じて複数の大学等が共同して、データキュレーター等を育成するシステムを検討し、推進する” (文部科学省科学技術・学術審議会学術分科会第 8 期学術情報委員会, 2016, p. 13) と述べられている。実際に、英国や米国では、助成機関による義務化されたのを契機として、データ公開の習慣や公開のためのリポジトリがない分野の研究者のために、図書館による支援が開始されている (Monastersky, 2013; Whyte & Sisu, 2014)。2.3.3 に示したように、日本においてもデータ公開支援の取り組みが開始されている。公開用リポジトリの提供などから支援を開始することが考えられる。

(2) 制度に関する課題 (業績／先取権)

データ公開を業績化することは、キャリアリスクの懸念への対策であると同時に、データ公開のインセンティブになると考えられている (Fecher et al., 2015; Ali-Khan, Harris, & Gold, 2017)。本研究においても、0 に示したように「現状ではデータ公開のインセンティブが充分

ではない」という指摘が 18 件あり、“公開した人は公開の労力が必要なのに対してメリットが無く、そのデータを利用する側にメリットがあるため、公開する側に大きなメリットがある仕組み作りが必要”といった記述がみられた。対照的に、論文の公開理由の自由記述では、OA 誌は査読期間が短い、IF が高いといった業績に関連するインセンティブが示されていた（表 3-9）。

データ公開を業績化して公開のインセンティブとすることは、国内外の政策でも検討や提言が行われている（Expert Advisory Group on Data Access, Economic and Social Research Council, Medical Research Council, Cancer Research UK, & Wellcome Trust, 2014; OECD, 2015; 国際的動向を踏まえたオープンサイエンスに関する検討会, 2015; 日本学術会議, 2016; G7 Science Ministers, 2017）。たとえば、2013 年 1 月に米国科学財団（National Science Foundation, NSF）は業績記入欄の名称を「Publication」から「Products」に変更して、申請者が公開したデータやプログラムコードの記載を可能にした（National Science Foundation [NSF], 2013）。データ公開が研究資金の獲得に繋がるならば、間接的に(1)の資金の問題への対応にもなると考えられる。また、データ公開の業績化は、データ公開が規範ではない分野への啓発の役割も果たすと考えられる。ただし、業績化に疑問を呈する自由記述もみられた（3.2.3）。データの取得、整備、公開のうち、どこまでを業績とするかについては慎重な議論が必要であろう。

データ公開を業績とするためには、公開データに対する一定の評価が必要になる。評価の副次的な効果として、公開データの利用におけるデータの品質の問題（表 3-16）への対策となり利用の促進が期待されること、自由回答にみられた悪意のある批判への懸念（3.2.4）が低減されることなどが考えられる。現状でも、分野によっては公開データの質が調査され（Peer, Green, & Stephenson, 2014; Roche, Kruuk, Lanfear, & Binning, 2015）、評価の枠組み（Costas, Meijer, Zahedi, & Wouters, 2013; Wuest, Mak-Dadanski, & Thoben, 2014）や標準（Zimmerman, 2008）、リポジトリによる基準⁴⁹⁾が検討されている。しかし、多くの分野では評価のためのプロトコルが確立されていない（Brase et al., 2009）。それでは、どのような方法でデータの評価を行えばよいのだろうか。

最も厳正な方法として、学術雑誌が専門家の査読によって掲載論文の質を保証しているように、データに対しても査読を行うことが提案されている（Parsons, Duerr, & Minster, 2010; Lawrence, Jones, Matthews, Pepler, & Callaghan, 2011; Costello, Michener, Gahegan, Zhang, & Bourne, 2013）。データの査読を経てデータ論文（data paper や data description）を掲載するデータジャーナル（南山, 2015）は、データの信頼性を保証する仕組みとして有効であると考えられる。また、一部の雑誌では補足資料やデータの査読を行っている⁵⁰⁾（Lindsay, 2017）。日

⁴⁹⁾ たとえば、次のものがある。“Quality Assurance”. UK Data Archive. <http://www.data-archive.ac.uk/create-manage/format/quality>, (accessed 2018-10-31).

⁵⁰⁾ たとえば、王立化学会の査読者用ガイドラインでは、全ての電子補足資料（Electronic Supplementary Information）を確認すること、通常は完全なチェックを求められない X 線結晶データについても充分な注意を払う必要があること、さらに専門家に確認を依頼することでもできるとしている。“Reviewer responsibilities: Reviewer procedure and policies”. Royal Society of Chemistry. <http://www.rsc.org/journals-books->

本の政策文書において公開データの質の担保や評価に関する記述がみられ、「学術情報のオープン化の推進について（審議まとめ）」では、“研究データは利活用可能な様式で公開され、かつその信頼性が確保されていることが重要である。研究データを基にした精度の高い成果の再現を可能とするためには、研究で使用したプログラムやソフトウェアのバージョン等についても、データ作成者が公開時に明示する必要がある”（文部科学省科学技術・学術審議会学術分科会第8期学術情報委員会, 2016, p. 8）と指摘されている。また、「内閣府報告書」では“研究データの技術的な品質の評価、またそれぞれの研究分野固有の基準に沿った評価について合意しておく必要がある。論文のピアレビューに相当するような評価の仕組みが必要”（国際的動向を踏まえたオープンサイエンスに関する検討会, 2015, p. 22）と述べられている。本研究による学術雑誌のポリシー調査では、54誌（全体の24.5%）がデータを査読する旨を明示していた（4.2.1）。しかし、研究者数や論文数の増大によって査読制度が限界を迎えつつあることが指摘されている（佐藤, 2016）現状において、さらに論文の根拠となるデータまで査読することは困難であろう。

表 3-18 に示したように、データの信頼性の判断基準として著者情報（70.9%）、研究手法の確かさ（62.8%）、そのデータを用いた論文（58.2%）が選択されていた。そこでデータを評価するための次善の策として、データを公開するリポジトリがデータの著者（作成者）の情報やデータを用いた論文への永続的かつ一意のリンクを確立することが考えられる。たとえば、ORCID⁵¹⁾による著者情報や、デジタルオブジェクト識別子（Digital Object Identifier, DOI）によるデータを用いた論文へのリンクなどを付与する仕組みを備えることが望ましいと考えられる。また、*Journal of Visualized Experiments*⁵²⁾は実験やデータの生成過程を動画によって確認可能にしている。オープンピアレビュー（Hachani, 2015）のように、データ公開者にデータ利用者からのフィードバックを掲載して他の研究者による評価を確認可能にすることなども業績化のための基盤となるのではないだろうか。

さらに簡便な評価の方法として、*Psychological Science* は論文とともにデータを公開した著者にバッジを付ける取り組みを行っており、公開率が3%未満から23%まで増加したと報告している（Kidwell et al., 2016）。業績化によるインセンティブの設計は、評価のコストと効果に対する検討が重要であると考えられる。

「先に論文を出版される可能性」は、データ公開経験の有無にかかわらず、問題視する回答者が多かった。また、Schmidt et al. (2016a) や Tenopir et al. (2015) と同様に、キャリア形成期である若年層の懸念が強かった（図 3-27）。所属別では大学の研究者が「問題である」を選ぶ比率が高く、キャリア形成期の大学の研究者による懸念が強いことが明らかになった。論文の出版前にデータ公開を求めることについては、本研究のみならず、2章に示した先行研

databases/journal-authors-reviewers/reviewer-responsibilities/, (accessed 2017-3-31).

⁵¹⁾ ORCID (Open Research and Contributor Identifier). <https://orcid.org>, (accessed 2018-10-31).

⁵²⁾ *Journal of Visualized Experiments*. <https://www.jove.com>, (accessed 2018-10-31).

究 (Fecher et al., 2015; Tenopir et al., 2015; Schmidt et al., 2016a) や環境学分野の調査 (Soranno, Cheruvelil, Elliott, & Montgomery, 2015) でもデータ公開の障壁となっていることが明らかにされている。生命科学分野など、1990 年代ごろから論文出版前にデータを公開するよう求めてきた分野もあるが (McCain, 1995), 同分野のように公開によって利点が生じる場合や、データジャーナル (林と村山, 2015) のように業績や引用、査読による質保証といったインセンティブが十分に提供される場合を除いて、論文出版前のデータ公開は受け入れ難いと考えられる。特別な事情がある場合を除いて、公開のタイミングは研究者が決定できることが望ましいといえよう。一方で、自由記述では、業績化や引用の徹底が保証されるならば、データ公開を前向きに捉えるという見解がみられた (3.2.3)。データ公開そのものが業績として認められるならば、また、データ引用が徹底されることによって先取権が確保されるならば、「先に論文を出版される可能性」への懸念は低減されると考えられる。

(3) 技術的解決が可能な課題 (機密・プライバシー／誤解・誤用)

薬学・医学分野のデータは、被験者のプライバシー情報や機密を含むため公開が難しいとされている。しかし、これらの分野の成果は市民の生命や健康に関わるため、データ公開による研究の加速や透明性の向上も重要である。米国国立衛生研究所 (National Institutes of Health, NIH) や EMA はデータ公開方針を公表しており (NIH, 2003; EMA, 2014), その効果も確認されている (Pham-Kanter et al., 2014)。世界中のがん患者と家族の生活の改善をめざす非営利イニシアティブである Project Data Sphere⁵³⁾のように、がん治療を目的として産学連携で個人情報や商業的な機密に配慮しながら臨床データと解析ツールを公開する取り組み (Bertagnolli et al., 2017) や、ブロックチェーン技術を応用してプライバシー情報を保護しながら医療データを公開する仕組みの開発 (Maxmen, 2018) も行われている。機密・プライバシー情報を含むデータは原則的に非公開としつつも、将来的には新たな手法を取り入れ、成果を生むことができれば、産学官の互恵的なデータ公開の機運が高まると期待できる。

「誤解や誤用の可能性」という質問文は、先行研究の“misinterpretation or misuse”を翻訳して用いており、意図的であるか否かという意味は含めなかった。しかし、3.2.4 に示した自由記述では、悪意のある誤用や改ざんなどの不正利用に対する強い懸念が複数示されていた。誤解や誤用への懸念がデータ公開の障壁となっているとする先行研究は多数あるが (Damvad, 2014; Fecher et al., 2015; Tenopir et al., 2015; Schmidt et al., 2016a), ネガティブかつセンシティブな性質の事項であるからか、実例の報告はみあたらなかった。実際に、このような事例があるのかどうかを明らかにすることを今後の課題としたい。また、(2)で言及したように、公開用のリポジトリがデータに紐づく論文や著者 (データの作成者) 情報への永続的なリンク

⁵³⁾ Project Data Sphere. <https://www.projectdatasphere.org/>, (accessed 2018-10-31).

を提供することによって、公開データの利用者のみならず、データを公開する研究者の誤解や誤用への懸念を低減できるのではないかと考えられる。

(4) 規範・慣習の醸成が必要な課題（引用／商用利用）

公開データの引用や被引用件数は、研究者にとっては公開のインセンティブになり (Fecher et al., 2015; Piwowar et al., 2007; Vandewalle, 2012; Henneken & Accomazzi, 2012; Dorch et al., 2015), 利用者や評価者にとっては公開データへのアクセスや判断の参考となる。2014 年 2 月に FORCE11⁵⁴⁾が公開したデータ引用原則の共同宣言 (Joint Declaration of Data Citation Principles, JDDCP) 最終版には、多くのデータアーカイブ、出版社、学協会、米国研究図書館協会 (Association of Research Libraries, ARL) といった組織が賛同している (Starr et al., 2015)。本調査の対象誌である *Atmospheric Chemistry and Physics* と *Geoscientific Model Development* (共に Copernicus GmbH 刊行) は FORCE11 のマニフェスト (Bourne et al., 2011) に賛同している旨を記載しており、*Global Biogeochemical Cycles* は FORCE11 と ESIP Commons のデータ引用ガイドライン (ESIP Commons, 2012) に準拠してデータセットを引用するよう奨励していた。また、ビッグ 3 (APA, シカゴ, MLA)⁵⁵⁾をはじめとする多くのスタイルガイドやデータリポジトリは、出版物と同様に、データやソフトウェアを参考文献リストに記載する場合の書式を示している。Dryad や PANGAEA など、データリポジトリがデータを利用した場合の引用スタイルを例示している場合も多い。

しかしながら現状では、公開されているデータやソフトウェアを再利用する際に、それらに関連する論文を引用することはあっても、データを直接引用している論文は少ない (Mooney, 2011; Mayo, Vision, & Hull, 2016; Robinson-García, Jiménez-Contreras, & Torres-Salinas, 2016)。2018 年 7 月 13 日の時点では、PMC に収録された 2017 年の論文 365,354 報のうち、データ引用を含む論文はわずか 850 報 (全体の 0.2%) であった⁵⁶⁾。

Borgman (2015) は、強制的な制度では研究者にデータ引用を浸透させることが難しく、現に実践されていないため、慣習や規範の醸成によらなければならないと論じている。出版社や学協会など、学術界のステークホルダーによる啓発活動が必要であると考えられる。たとえば、データの利用規約などで論文の引用を促すことによって、引用せずに利用される可能性への懸念を低減できるかもしれない。また、論文の引用と同様に、データ引用が一般化すれば、データ公開理由の 1 位であった研究成果の認知向上 (Starr et al., 2015) にも繋がると考

⁵⁴⁾ 科学的知識の共有を推進する国際イニシアティブ。FORCE11 (Future of Research Communications and e-Scholarship). <https://www.force11.org>, (accessed 2018-10-31).

⁵⁵⁾ 学術文献の代表的なスタイルマニュアル。APA (American Psychological Association) style, シカゴ (The Chicago Manual of Style), MLA (Modern Language Association) style を指す。

⁵⁶⁾ 2018 年 4 月、PMC と PubMed は、検索フィルターに「データ引用を含む」などを追加した。2017 年に出版されたデータ引用を含む論文の検索式は、次の通りである。(("2017/01/01"[Publication Date] : "2017/12/31"[Publication Date])) AND has data citations[filter]

えられる。

引用方法を標準化することによってデータの再利用状況の追跡や評価を可能にすることも期待されている (Starr et al., 2015)。データが直接引用されることによって、データは出版物と同様の研究成果である (Martone, 2014) という認識も浸透するのではないだろうか。データを公開している論文は、そうでない論文よりもよく引用されるという複数の研究結果や報告がある (Gleditsc, Metelits, & Strand, 2003; Piwowar et al., 2007; Bruna, 2010; Sears, 2011; Vandewalle, 2012; Henneken & Accomazzi, 2012; Piwowar & Vision, 2013; Dorch et al., 2015)。McKiernan et al. (2016) のレビュー論文では、オープンサイエンス⁵⁷⁾は引用だけではなく、メディア報道、新たな共同研究、キャリアアップ (career advancement)、助成金獲得などに影響を与えることが示されている。こうした研究のビジビリティの向上による利点が調査され、広まることによって、データ公開のインセンティブとなるのではないだろうか。

商用利用について、オープンサイエンス政策では国益や知的財産に留意しつつ、“企業や個人においては、科学的成果を活用・再利用して新しい製品や新しいサービス（市場）を生み出すこと” (国際的動向を踏まえたオープンサイエンスに関する検討会, 2015, p.1) が期待されている。2017 年 6 月、*Scientific Data* は、同誌に投稿したデータにクリエイティブ・コモンズ・ライセンス⁵⁸⁾の CC BY-NC (非営利) の適用を認めない、すなわち商用利用の制限を認めないとする方針を発表した (Scientific Data, 2017)。大学や公的機関・団体の研究者であつても商用利用への懸念が強かったが、研究成果が企業のイノベーションに一定の役割を果たしている (池内, 本橋, 田村, 塚田, 2017) ことから、商用利用に関する議論とコンセンサスの形成が必要であると考えられる。

[2] データ公開経験や態度と関連がある経験・認識

データ公開経験や態度と関連がある経験・認識について、「自発的にデータ公開を行っている分野」(植物・動物学とコンピュータサイエンス) は、(a)データの提供頻度、(b)被提供頻度、(c)公開データの入手経験、(c-2)再利用・再分析する頻度が高かった。一方、「データ公開が困難な分野」(臨床医学と精神医学・心理学) は、(a)データの提供頻度、(b)被提供頻度、(c)公開データの入手経験がある研究者の比率が低く、(c-2)再利用・再分析を行っている回答者は両分野とも 0%であった。ただし、非公開理由が解決した場合の公開意思は、平均よりも高か

⁵⁷⁾ McKiernan らは、OA、オープンデータ、オープンソース、オープンスカラシップを列挙した上で、オープンピアレビューなども取り上げている。

⁵⁸⁾ 国際非営利組織クリエイティブ・コモンズが提供する、著作物にライセンスを表示するためのツール。日本を含む多くの国ではデータに著作権が認められていないが、知名度が高いことから、公開データのライセンスの意思表示に使用される場合が多い。クリエイティブ・コモンズ・ライセンスとは、Creative Commons Japan. <https://creativecommons.jp/licenses/>, (accessed 2018-10-31).

った。

3.3.8 に示した「データ公開経験をもつ研究者」、3.3.9 に示した「データ公開に消極的な研究者」、および以上の結果を表 4-29 に示す。「+」は該当者は経験がある比率や頻度が高いことを、「-」は該当者は経験がある比率や頻度が低いことを表している。

表 4-29 データ公開経験・態度と関連がみられた経験・認識

質問項目	データ公開	消極的	自発的	困難
データの提供頻度	+	-	+	-
データの被提供頻度	+	-	+	-
公開データの入手経験	+	-	+	-
研究の参考	+			
再利用・再分析	+	-	+	-
再現・追試	+			
公開データの利用希望	+	-		
OA 経験	+			
異分野の研究者による理解	+			

「データ公開」=データ公開経験をもつ研究者、「消極的」=データ公開に消極的な研究者、「自発的」=自発的にデータを公開している分野、「困難」=データ公開が困難な分野

先行研究では、個人的あるいは研究グループ内でのデータ共有（提供や被提供）経験がデータ公開に繋がる、あるいは、データの再利用経験がデータ公開経験に繋がると論じられていた（Piwowar, 2011; Damvad, 2014; Berghmans et al., 2017）。本研究では、データの提供頻度、データの被提供頻度、公開データの入手経験、入手したデータの再利用・再分析経験と、データ公開経験に相関がみられた。本研究は、これらの経験とデータ公開経験の因果関係までは明らかにできなかったが、関連がある可能性が示唆された。データの入手経験や再利用経験がデータ公開に繋がると仮定するならば、の[1]に示したりポジトリ整備がいっそう重要であると考えられる。

[3] データ公開の主要な課題に対する方策と期待される効果

以上のまとめとして、図 4-5 に日本の研究者の認識に基づくデータ公開の主要な課題、およびステークホルダーによる対策と期待される効果を示す。図では、「対策」は網掛けした四角形、それによって「期待される効果やインセンティブ」は実線の角が丸い四角形、対策または効果によって低減されると考えられる「主要な課題」は破線の角が丸い四角形で示した。ここまでみてきたように、「対策」、「期待される効果やインセンティブ」、「主要な課題」は相互に関連している。たとえば、データ公開リポジトリを整備し、データを FAIR 原則に準拠し

て、Findable (見つけられる), Accessible (アクセスできる), Interoperable (相互運用できる), Re-usable (再利用できる) 状態で公開可能とする。具体的には ORCID で著者情報へ、DOI で論文へのリンクを確立しておくことによって、評価や査読、およびデータを利用する際に信頼性を確認するためのアクセスが可能になる。データへの評価や査読が行われると、誤解や誤用への懸念が低減し、データの信頼性を増すことができる。さらに業績化して昇進や助成金獲得に繋がるようになれば資金の問題が低減したり、先に論文を出版される可能性への懸念が低減したりする。また、リポジトリにデータの引用方法を明示しておくことによって、データ引用が促進されれば、データの認知度の向上やデータ利用の増加につながり、先に論文を出版される可能性や引用せずに利用される可能性といったキャリアリスクへの懸念が低減されると考えられる。

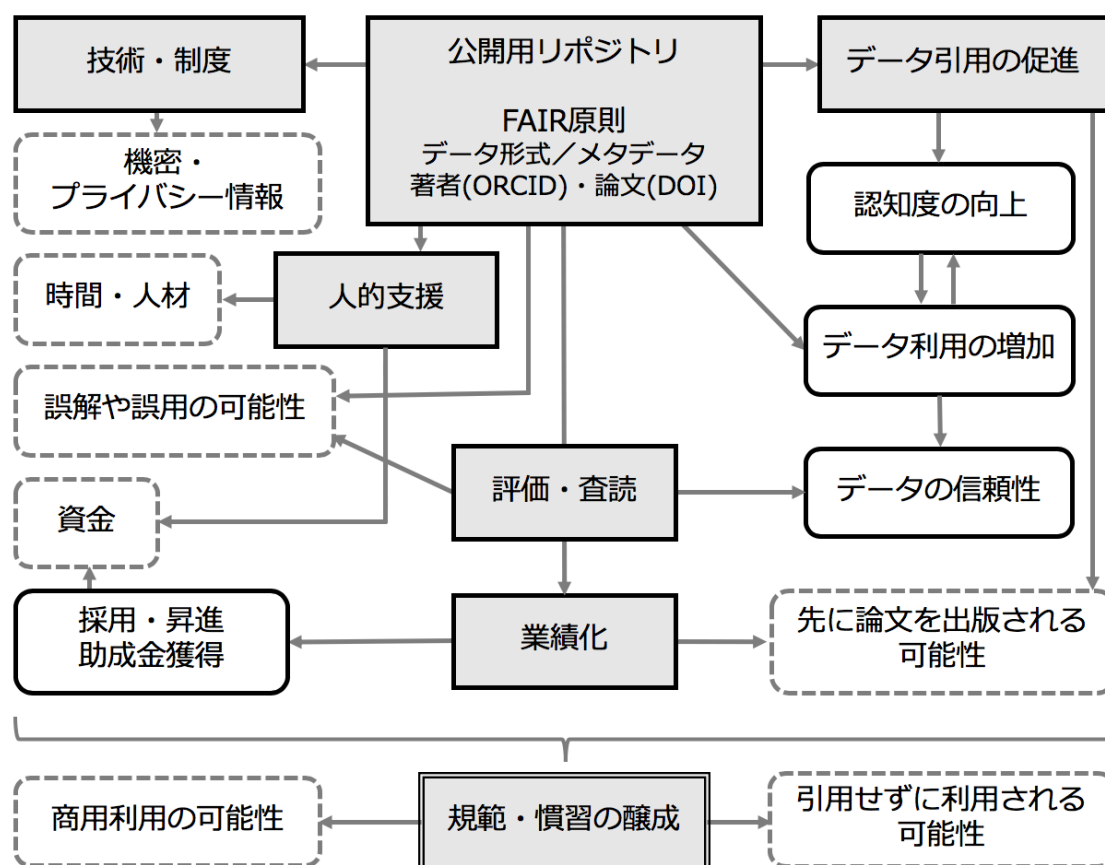


図 4-5 日本におけるデータ公開の主要な課題と対策、期待される効果

データの利用と公開の懸念が低減し、効果やインセンティブが認識され、データの公開と利活用のサイクルが循環するようになれば、データ公開の規範や慣習が醸成されるのではないだろうか。規範や慣習の拡がりによって、徐々に学術コミュニティにおける商用利用やデータ引用に関する認識が変化していくのではないかと考えられる。

4.4 まとめ

学術雑誌によるデータ公開要求の状況を明らかにすること、および日本の研究者によるデータ公開経験と比較することを目的として、データ公開ポリシーを調査した。調査対象として、ESIの雑誌リストを用いて22分野のIFの高い原著論文誌を各10誌ずつ、合計220誌を選択した。2014年4月から6月にかけて、調査対象誌のポリシー（投稿規定などデータに関する記述）をウェブサイトで調査した。ポリシーは、RP（リポジトリにデータを公開して論文に識別子を記すポリシー）と、SP（補足資料にデータを掲載するポリシー）を調査して、それぞれの強度を「必須」、「推奨」、「受諾」、「なし」の4段階に分類した。また、Ulrichsweb.comなどを用いて雑誌の特徴を調査した。その上で、データ公開ポリシーの強度と日本の研究者によるデータ公開率を分野ごとに比較した。

調査対象である220誌のうち、何らかのデータ公開ポリシーを掲載していた、すなわちRPかSPのどちらか、あるいは両方を掲載していた雑誌は204誌（92.7%）であった。RPの掲載率は59.5%、SPの掲載率は89.5%であった。RPは要求の強度が高い「必須」と「推奨」が多く（合計59.1%）、SPは要求の強度が低い「受諾」が多かった（62.3%）。

RPとSPの掲載率や強度は分野によってばらつきがみられた。データ公開が盛んな生物・医学分野であっても下位の研究領域では差があり、農業科学、薬理学・毒性学、精神医学・心理学は掲載率や要求の強度が中～低度であった。データの再利用が盛んな地球科学、宇宙科学、社会学であっても、掲載率や要求の強度は中程度であり、再分析したデータの公開を求める雑誌は多くなかった。また、工学と経済学・経営学は、RPを全く掲載していなかった。

調査対象誌の特徴として、(1)倫理やCOIの開示に関する記述の有無、(2)出版者（学協会と商業出版社）、(3)OAステータスを調べた。(1)倫理やCOIの開示に関する記述がある雑誌は、ポリシーの掲載率や要求の強度が高いことがわかった。これは、本研究独自の調査項目であり、データ公開ポリシーは研究不正の抑制に関連がある可能性が示唆された。(2)出版者は、先行研究では学協会の掲載率が高いという結果が示されていたが、本調査では商業出版社の方が掲載率も要求の強度も高かった。なお、(1)と(2)については分野によるばらつきが大きかったため、生物・医学グループと自然科学グループにわけて確認した。その結果、分野によるばらつきを勘案しても、ほぼ同じ結果であった。(3)OAステータスについて、OA誌は全てポリシーを掲載しており、RPは掲載率も強度もOA誌、制約つきOA誌、非OA誌の順に高かった。一方、SPについては、掲載率は制約つきOA誌（92.8%）、非OA誌（82.5%）、OA誌（73.3%）の順に高く、「必須」の比率は非OA誌（23.8%）、制約つきOA誌（8.8%）、OA誌（6.7%）の順であった。オープンサイエンス政策では、研究成果へのパブリックアクセスの観点から論文のOAとデータ公開を推進しているが、雑誌のポリシーについては、OA誌の方がデータ公開を強く要求しているとまではいえなかった。

予備調査、および本調査でポリシーを調査した過程で得られた知見から、雑誌のポリシー

の課題として、(1)定形ポリシーの策定と(2)データの公開先リポジトリの指定について検討した。(1)定形ポリシーは複数の雑誌で採用されており、統一ポリシーの策定も進められている。本研究の調査対象とした COPE ガイドラインなどを採用するように、定形ポリシーが広く普及すれば、出版者にとっても研究者にとっても便益が大きいと考えられる。(2)データの公開先リポジトリが指定されていない場合、適切な公開先がなく、または、適切な公開先がよくわからず、時間が不足している研究者にとって投稿の障壁となる可能性がある。適切なリポジトリやリポジトリのリストを指定することで、こうした問題を低減することができるのではないかと考えられる。

分野別に RP の掲載率と 3 章で調査したデータの公開率を比較した結果、両者には正の相関がみられた。つまり、RP の強度が高い分野ほどデータ公開率が高いことがわかった。一方で、RP の強度と比較したデータ公開率に乖離がある分野もみられた。RP に対して公開率が高く、雑誌のポリシーによらず自発的にデータを公開していると考えられたのは植物・動物学とコンピュータサイエンスであった。3 章で明らかにした内的要因を確認したところ、この 2 分野に共通していたのは、データの公開理由として「科学研究や成果実装の推進」の選択率が高いこと、データの被提供頻度が高いこと、商用利用される可能性や機密・プライバシー情報への懸念が低いことであった。公開率が低くデータ公開が困難であると考えられたのは臨床医学と精神医学・心理学であった。臨床医学はデータの提供頻度や入手経験が低く、再分析・再利用を行う回答者は全くいなかった。データに機密情報を含む回答者の比率が高く、機密・プライバシー情報への懸念も高かった。その一方で、雑誌や助成機関（日本医療研究開発機構 [AMED], 2018）によってデータ公開が求められているため、課題解決の必要性が高いと考えられる。精神医学・心理学は RP の強度が低く、リポジトリの充足度やデータの提供・被提供、入手や利用の頻度も低かった。また、データに機密・プライバシー情報を含む回答者の比率が高かった。ただし、データ公開意思がある回答者の比率が比較的高く、データを利用したいとする他分野の回答者の比率も高かった。

3 章ではデータ公開に関する障壁、すなわち懸念や資源の不足感はいずれも高いことが明らかになった。そこでデータ公開経験がある研究者や自発的にデータ公開を行っている分野の研究者にとって低い障壁、あるいはデータ公開に消極的な研究者やデータ公開が困難な分野の研究者にとって高い障壁などに基づき 11 の主要な課題を抽出して、(1)資源に関する課題（公開用リポジトリ、公開のための時間、人材、資金）、(2)制度に関する課題（業績、先取権）、(3)技術的解決が可能な課題（機密・プライバシー情報、誤解や誤用の可能性）、(4)規範・慣習の醸成が必要な課題（引用、商用利用）に整理した。また、データ公開経験やデータ公開に対する態度はデータの提供経験や再利用・再分析経験と関連があることがわかった。

5 章 おわりに

本研究は、日本の研究者によるデータ公開状況およびデータ公開に対する認識を明らかにすることを目的として、質問紙調査と学術雑誌のデータ公開ポリシーの調査を行った。本章では、3つの研究課題の結果をそれぞれ示した上で研究の結論を述べる。また、研究の新規性、応用、限界、今後の課題について述べる。

5.1 主要な結果と結論

5.1.1 研究課題と結果

[1] 日本の研究者によるデータ公開の状況

[RQ1]日本の研究者はどの程度データを公開しているのかを明らかにするために、2016年11月から12月にかけて、文部科学省科学技術・学術政策研究所（NISTEP）科学技術動向研究センターが運営する科学技術専門家ネットワーク（1,983名）を対象として、ウェブアンケートシステムを用いた調査を行った。大学、公的機関・団体、企業の研究者1,398名（有効回答率70.5%）の回答を分析した結果、データ公開の経験をもつ回答者の割合（以下、「データ公開率」）は、51.0%であった。オープンアクセス（Open Access, OA）の経験をもつ回答者の割合は70.9%であり、データ公開率の方が約20ポイント低かった。データ公開率を所属機関別にみると、大学（56.9%）、公的機関・団体（55.6%）、企業（32.4%）の順に高かった。

2014年のWiley社による調査では52%（Ferguson, 2014）、2016年のCWTS/Elsevierによる調査では66%（Berghmans et al., 2017）であったことから、日本の研究者によるデータ公開率は国外の研究者と比較するとやや低いと考えられる。国外の研究者を対象とした先行研究では、助成機関のポリシー（Wallis, 2013; Fecher et al., 2015; Kim & Burns, 2015; Schmidt et al., 2016a）や所属機関のポリシー（Fecher et al., 2015; Kim & Stanton, 2015; Schmidt et al., 2016a）がデータ公開の主要な動機となっていたが、本研究ではいずれも選択率が低く、助成機関のポリシーを選択した回答者は6.3%、所属機関のポリシーを選択した回答者は23.7%であった。調査時点の2016年には、助成機関や研究機関がほとんどデータ公開を要求していなかったことが日本の研究者のデータ公開率が低い要因の1つであると考えられる。

しかし本調査を実施した後、2017年に科学技術振興機構（JST）（2017）が「オープンサイエンス促進に向けた研究成果の取扱いに関するJSTの基本方針」と「運用ガイドライン」を公開した。次いで2018年には新エネルギー・産業技術総合開発機構（NEDO）（2018）が「NEDO

プロジェクトにおけるデータマネジメントに係る基本方針」を、日本医療研究開発機構（AMED）（2018）が「データマネジメントプランの提出について」を公開した。国外の要求に比べると緩やかではあるものの、日本においても助成機関によるデータ公開の要求が始まりつつある。また、研究機関によるデータ公開方針の策定を促すものとして、内閣府の国際動向を踏まえたオープンサイエンスの推進に関する検討会（2018）が「国立研究開発法人におけるデータポリシー策定のためのガイドライン」を公開した。こうしたデータ公開に関する制度がデータ公開率の上昇に一定の影響を与える可能性があると考えられる。

[2] 日本の研究者によるデータ公開の内的要因

[RQ2]日本の研究者によるデータ公開の内的要因は何かを明らかにするために、質問紙調査でデータ公開の動機と障壁について尋ね、分野や属性による差を分析した。

日本の研究者にとってのデータ公開の動機について、データを公開する理由の1位は「研究成果の認知度向上」（58.5%）、2位は「雑誌のポリシー」（43.8%）であった。先行研究で公開の動機とされていた「分野の規範」は分野別にみても選択率が低く、高い分野であっても数学（28.6%）、植物・動物学（25.0%）、分子生物学・遺伝学（22.6%）であった。また科学的利他性に関する項目、すなわち「科学研究や成果実装を推進したいから」（26.1%）、「他の研究者からのリクエストに応じて」（25.4%）、「オープンデータに貢献したいから」（10.5%）も高いとは言えなかった。

データ公開の障壁について、データの非公開理由の1位は「雑誌のポリシーではないこと」（26.4%）、2位は「時間がないこと」（25.9%）であった。1位～5位の選択率はいずれも20%台であり（3位「所属機関にポリシーがないから」（22.4%））、4位「ニーズがないと思うから」（22.1%）、5位「業績にならないから」（21.7%））、突出した理由はみられなかった。論文の非公開理由は、1位「雑誌のポリシーではないこと」（60.4%）と2位「公開のための資金」（39.6%）に集中しており、かつ、解決すれば公開して良いと考える回答者が78.1%にのぼったのに対して、データは28.4%にとどまった。また、データ公開につながると考えられる非公開理由を特定するために、データ公開意思をもつ回答者群ともしない回答者群とで選択率を比較したが、差のある項目はなかった。ただし、データ公開の業績化や評価、公開データの引用が保証されるならばデータ公開に前向きになるとの意見もみられた。自由記述では、純粹数学など公開するデータがない分野の存在が確認された。

データを管理・公開するための資源については、「人材」、「時間」、「資金」の充足度が低かった。ここでは所属機関による差がみられ、いずれの資源も大学、公的機関・団体、企業の順に充足度が低かった。また、分野別のデータ公開率と公開用リポジトリの充足度には正の相関がみられた。

データを公開する場合の懸念は「引用せずに利用される可能性」と「先に論文を出版される可能性」（以下、「キャリアリスク」）が強かった。「引用せずに利用される可能性」につい

て、自由記述では“盗用”や“剽窃”といった意図的な利用を示唆する表現がみられた。「先に論文を出版される可能性」については、論文の出版前に公開することに対する強い抵抗がうかがえ、“自殺行為”といった表現もみられた。属性別にみると、若年層や大学に所属する研究者は、キャリアリスクに対する懸念が特に強かった。一方、企業の研究者は「機密・プライバシー情報」や「商用利用される可能性」への懸念が強く、特許との関わりを指摘する回答も17件みられた。分野別にみると、臨床医学、免疫学の研究者は「機密・プライバシー情報」や「商用利用される可能性」への懸念が強かった。

データの公開経験がある日本の研究者は、(a)個人的なデータの提供頻度、(b)個人的なデータの被提供頻度、(c)公開データの入手経験、および利用頻度((c-1)自身の研究の参考にする、(c-2)再利用・再分析する、(c-3)結果を再現・追試する)が高く、(d)公開データの利用希望、(e)OAの経験を有する回答者の比率も高いことがわかった。また、データを公開した場合、異分野の研究者による理解が容易であると考えている回答者群は、データ公開率が高かった。一方、データ公開経験と扱っているデータの種類、すなわち、データ量、機密情報の有無、データの所有権の有無には差が認められなかった。

データ公開の経験がない研究者のうち、現在の非公開理由が解決したとしても公開しないと回答した研究者（以下、「データ公開に消極的な研究者」）は、(a)データの提供頻度、(b)被提供頻度、(c)公開データの入手経験や(c-2)再利用・再分析の頻度が低く、(d)公開データの利用を希望する回答者も少なかった。

[3] データ公開の外的要因の状況と日本の研究者によるデータ公開状況

[RQ3]データ公開の外的要因はどのような状況にあるのか、この外的要因と日本の研究者によるデータ公開状況に乖離はあるのかどうかを明らかにするために、学術雑誌のデータ公開ポリシーを分野横断で調査し、質問紙調査の結果と比較した。

学術雑誌のデータ公開ポリシーを調査した結果、調査対象である220誌のうち、何らかのデータ公開ポリシーを掲載していた、すなわちリポジトリにデータを公開して論文に識別子を記すポリシー（Repository Policy, RP）または補足資料にデータを掲載するポリシー（Supplemental Materials' Policy, SP）のどちらか、あるいは両方を掲載していた雑誌は204誌（92.7%）であった。RPの掲載率は59.5%、SPの掲載率は89.5%であった。RPは要求の強度が高い「必須」と「推奨」が多く（合計59.1%）、SPは要求の強度が低い「受諾」が多かった（62.3%）。

分野別に確認すると、生物学と医学の関連領域はRPの掲載率や強度が高い傾向にあり、定評のある生物データのリポジトリへの公開を求めている。しかし、農業科学は10誌のうち7誌、薬理学・毒性学は6誌、精神医学・心理学は2誌であり、領域内で差がみられた。データ共有による二次分析が盛んな地球科学は7誌、宇宙科学は6誌、社会科学は4誌であり、RPの掲載率は中程度であった。研究内容によってデータが用いられていると推測される分野

であっても、コンピュータサイエンスは2誌、数学は1誌にしか掲載されておらず、工学と経済学・経営学は全く掲載されていなかった。

調査対象誌の特徴として、(1)倫理や利益相反 (COI) の開示に関する記述の有無、(2)出版者 (学協会と商業出版社)、(3)OA ステータスを調べた。(1)倫理や COI の開示に関する記述がある雑誌は、ポリシーの掲載率や要求の強度が高いことがわかった。(2)出版者は、先行研究では学協会の掲載率が高いという結果が示されていたが、本調査では商業出版社の方が掲載率も要求の強度も高かった。なお、(1)と(2)については分野によるばらつきが大きかったため、生物・医学グループと自然科学グループにわけて確認したが、分野によるばらつきを勘案しても、ほぼ同じ結果であった。(3)OA ステータスについて、OA 誌は全てポリシーを掲載しており、RP は掲載率も強度も OA 誌、制約つき OA 誌、非 OA 誌の順に高かった。一方、SP については、掲載率は制約つき OA 誌 (92.8%)、非 OA 誌 (82.5%)、OA 誌 (73.3%) の順に高く、「必須」の比率は非 OA 誌 (23.8%)、制約つき OA 誌 (8.8%)、OA 誌 (6.7%) の順であった。オープンサイエンス政策では、研究成果へのパブリックアクセスの観点から論文の OA とデータ公開を推進しているが、雑誌のポリシーについては、OA 誌の方がデータ公開を強く要求しているとまでは言えなかった。

データ公開要求に対して日本の研究者はどの程度データを公開しているのかを確かめるため、分野別に RP の掲載率とデータの公開率を比較した結果、両者には正の相関がみられた。つまり、RP の掲載率が高い分野ほどデータ公開率が高いことがわかった。一方で、RP の掲載率と比較したデータ公開率に乖離がある分野もみられた。RP の掲載率に対してデータ公開率が高く、雑誌のポリシーによらず自発的にデータを公開していると考えられたのは植物・動物学とコンピュータサイエンスであった。この2分野に共通していたのは、データの公開理由として「科学研究や成果実装の推進」の選択率が高いこと、データの被提供頻度が高いこと、商用利用される可能性や機密・プライバシー情報への懸念が低いことであった。RP の掲載率に対してデータ公開率が低くデータ公開が困難であると考えられたのは、臨床医学と精神医学・心理学であった。臨床医学はデータの提供頻度や入手経験が低く、再分析・再利用を行う回答者は全くいなかった。データに機密情報を含む回答者の比率が高く、機密・プライバシー情報への懸念も高かった。精神医学・心理学は RP の掲載率が低く、リポジトリの充足度やデータの提供・被提供、入手や利用の頻度も低かった。また、データに機密・プライバシー情報を含む回答者の比率が高かった。ただし、データ公開意思がある回答者の比率が比較的高く、データを利用したいとする他分野の回答者の比率も高かった。

5.1.2 結論

本研究は、日本の研究者を対象として、データ公開の実践状況と認識を明らかにするために質問紙調査を実施し、学術雑誌によるデータ公開ポリシーの調査の結果と比較した。学術雑誌の要求と日本の研究者によるデータの公開状況を分野別に比較した結果から、自発的に

データ公開を行っている分野とデータ公開が困難な分野があることがわかった（図 5-1）。また，研究全体を通じて主要な課題，およびデータ公開が適さない場合があることも示唆された。

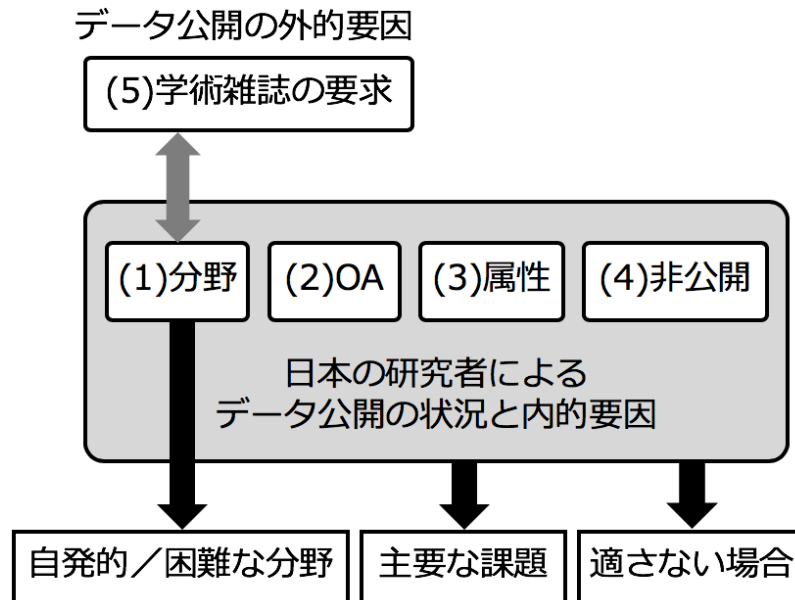


図 5-1 本研究の比較項目と新たな知見

日本の研究者によるデータ公開の実践状況，すなわちデータ公開率（51.0%）は，国外の研究者と比較するとやや低かった。助成機関や研究機関によるデータ公開の義務化がほとんど行われていないことが主要な要因の1つであると考えられる。

分野別の公開率は，最も高い植物・動物学（84.2%）から最も低い精神医学・心理学（30.8%）まで分野間の差がみられた。しかしデータの公開理由のうち「分野・コミュニティの規範」は選択率が低く，上位の分野であっても3割に満たなかった。一方，データ公開用リポジトリの充足度については分野による差があり，充足度が高い分野ほど公開率も高かった。

論文はデータよりも公開率が高く（70.9%），分野による差がみられたが，データ公開との関連はみられなかった。つまり，論文の公開率が高い分野であっても，データの公開率が高いとは限らなかった。OAは雑誌のポリシー（60.4%）と資金（39.6%）が主な非公開理由であり，解決した場合に公開意思がある回答者は78.1%にのぼったが，データは3割以上の回答者が選択した非公開理由はなく，解決したとしても公開する意思がある回答者は28.4%であった。データ公開はOAと比較して実践されておらず，成熟度が低いといった指摘を支持する結果であった（OECD, 2015）。

所属別では，大学（56.9%），公的機関・団体（55.6%），企業（32.4%）の順に公開率が高かった。大学の研究者は人材，時間，資金の充足度が低く，キャリアリスクへの懸念が強かった。企業の研究者は機密・プライバシー情報や商業利用される可能性への懸念が強かった。

年齢層別では、先行研究と同様に若年層ほどキャリアリスクへの懸念が強かった。(Tenopir et al., 2015; Schmidt et al., 2016a)。

データ公開経験のある回答者群は、経験のない回答者群と比較して、公開用のリポジトリ、時間、人材、資金の充足度が高く、商用利用される可能性や機密・プライバシー情報への懸念は弱かった。また、データの共有や入手、利用の頻度は高く、公開データの利用希望も多かった。データ公開経験のない回答者群のうち、公開意思のある回答者は公開意思のない回答者と比較して、先に論文を出版される可能性や商用利用される可能性への懸念が弱く、データ共有や入手、利用の頻度も高かった。なお、資源の充足度には差がみられなかった。

学術雑誌のデータ公開要求について、RP と SP を掲載している雑誌は 92.7%であった。RP は 59.5%が掲載しており、59.1%は要求の強度が高い「必須」または「推奨」であった。SP は 89.5%が掲載していたが、62.3%は要求の強度が低い「受諾」であった。分野別に確認すると、生物学や医学の関連領域は RP の掲載率や強度が高い傾向にあり、定評のある生物データのリポジトリへの公開を求めている。しかし、農業科学は 10 誌のうち 7 誌、薬理学・毒性学は 6 誌、精神医学・心理学は 2 誌であり、領域内で差がみられた。データ共有による二次分析が盛んな地球科学は 7 誌、宇宙科学は 6 誌、社会科学は 4 誌であり、RP の掲載率は中程度であった。研究内容によってデータが用いられていると推測される分野であっても、コンピュータサイエンスは 2 誌、数学は 1 誌にしか掲載されておらず、工学と経済学・経営学は全く掲載されていなかった。

RP の掲載率と日本の研究者のデータ公開率を分野別に比較した結果、両者には正の相関がみられ、RP の掲載率が高い分野ほど公開率が高いことが明らかになった。生物学や医学の関連領域は RP の掲載率と同様にデータ公開率も高い傾向にあったが、農業科学 (56.5%) と薬理学・毒性学 (51.4%) は中程度であり、精神医学・心理学 (30.8%) は低く、領域内で差がみられた。また、データの二次分析が行われてきた分野であっても、RP の強度と同様にデータ公開率は中程度であり、地球科学は 52.4%、宇宙科学は 50.0%、社会科学は 50.0%であった。

また、RP の掲載率とデータ公開率を比較して、乖離がみられる分野 ($\pm 15\%$ の範囲からの外れ値)を確認したところ、RP の掲載率に対してデータ公開率が高かったのは植物・動物学 (84.2%) とコンピュータサイエンス (65.1%) であった。この 2 分野は、雑誌の要求によらず、内的要因によって自発的にデータ公開を進めていると推測され、(a)データの提供頻度、(b)被提供頻度、(c)公開データの入手経験、(c-2)再利用・再分析する頻度が高かった。一方、RP の強度に対してデータ公開率が低かったのは臨床医学 (45.5%) と精神医学・心理学 (30.8%) であった。この 2 分野は、データ公開が困難であると推測され、(a)データの提供頻度、(b)被提供頻度、(c)公開データの入手経験がある研究者の比率が低く、(c-2)再利用・再分析を行っている回答者は両分野とも 0%であった。ただし、非公開理由が解決した場合の公開意思は、平均よりも高かった。

データ公開に関するさまざまな障壁が明らかになったが、「これを解決すればデータ公開が進む」と判断できるような単一の課題は特定されなかった。つまり、現状ではデータ公開に関する課題は多く、データ公開自体が緒に就いた段階であると考えられる。そこで、ここまでの分析から課題の優先順位を検討するために、(A)データの公開理由のうち選択率の低い項目、(B)データの非公開理由のうち選択率の高い項目、(C)データ公開の障壁のうち選択率の高い項目、(D)データ公開経験がある研究者にとって低い障壁、(E)データ公開に消極的な研究者にとって高い障壁、(F)自発的にデータ公開を行っている分野の研究者にとって低い障壁、(G)データ公開が困難な分野の研究者にとって高い障壁のうち、2項目以上あてはまる要因を主要な課題とみなすこととした。その結果、主要な課題として11の要因が抽出された。学術機関、学術基盤の提供者、学術出版社、助成機関、政府関係者といったステークホルダーが対策を講じる観点から整理して、(1)資源に関する課題（公開用リポジトリ、公開のための時間、人材、資金）、(2)制度に関する課題（業績、先取権）、(3)技術的解決が可能な課題（機密・プライバシー情報、誤解や誤用の可能性）、(4)規範・慣習の醸成が必要な課題（引用、商用利用）に分類した。これらの課題は相互に関連しており、1つの対策が複数の効果やインセンティブとなる可能性や、複数の対策によって1つの懸念を低減する可能性があると考えられる。

また、調査を通じてデータ公開に適さない場合があることが示唆された。分野では、純粋数学のようにデータを扱わない分野、臨床医学、免疫学など機密・プライバシー情報や商業利用への懸念が強い分野、工学のように製品や企業秘密を含む分野などが該当する。また、物性物理学は他の研究者が理解し再利用することが困難であるとの指摘もみられた。企業の研究者は機密・プライバシー情報と商用利用される可能性への懸念が特に強く、特許のため公開は困難であることが明らかになった。データ公開に係る制度設計においては一律の公開を求めないことや、こうした分野の研究者が評価などで不利益を被ることがないように留意する必要があると考えられる。

5.2 研究の新規性

本研究の新規性として、これまで十分に調査されてこなかった点や新たな知見のうち、主要な6点について述べる。表 5-1 に6項目と先行研究の状況を示す。

表 5-1 本研究の新規性と先行研究の状況

本研究	先行研究
1. 日本の研究者を対象としたデータ公開の実態調査	小野ら, 2016 (38 名)
2. 同一の被験者を対象としたデータ公開と OA の比較調査	未実施
3. 研究者の所属機関（大学と企業）による相違点の比較	未実施
4. 分野を横断したデータ公開ポリシーの調査	未実施
5. データ公開ポリシーと研究不正対策の関連	未実施
6. データ公開ポリシーの強度と実際のデータ公開状況の比較	Vines et al., 2013b (12 誌)

[1] 日本の研究者を対象としたデータ公開の実態調査

日本の研究者を対象としたデータ公開の実態は、小野ら、(2016) による 38 名を対象とした調査、および Wiley 社の調査に日本の研究者によるデータの公開率が示されていた (Ferguson, 2014) 他には、ほとんど調査が行われてこなかった。本研究は多分野から 1,398 名の回答を得たことによって、国外調査の結果と比較することができた。たとえば、日本の研究者はデータ公開経験をもつ比率がやや低いこと、データ公開の理由として科学的利他性をあまり選択しないことなどがわかった。

[2] 同一の被験者を対象としたデータ公開と OA の比較調査

これまでデータ公開に関する実態調査も OA に関する実態調査も多数行われてきたが、同一の研究者に、同一の質問項目でデータ公開と OA の実践状況や認識を尋ねる調査は管見の限り行われていない。オープンサイエンス政策では、公的資金による研究成果として出版物とデータへのオープンなアクセスを求めているが、実践状況、公開理由、非公開理由、公開意思が異なることがわかった。たとえば、いずれも雑誌のポリシーではないことが非公開理由として多く選ばれていたが、非公開理由が解決した場合に論文は「公開する」という回答者が多かったのに対してデータは「公開しない」、または「わからない」とする回答者が多かった。オープンサイエンス政策を進めるのにあたって、データ公開には特有の懸念があることや、資源の状況に留意する必要があると考えられる。

[3] 研究者の所属機関（大学と企業）による相違点の比較

先行研究では、調査対象者に企業の研究者が含まれている場合もあったが (Tenopir et al., 2011; Schmidt et al., 2016a), 企業の研究者の比率が少なく、比較分析は行われてこなかった。

本研究では、企業の研究者を調査対象に含み、大学や公的機関・団体の研究者と比較した結果、企業の研究者は商用利用される可能性や機密・プライバシー情報への懸念が強く、本人以外にカレントデータの所有権をもつ人・組織やカレントデータに機密情報を含む比率も高いことがわかった。

内閣府の知的財産戦略本部は、民間投資によるデータと人工知能の利活用を目指すため、営業秘密に関するデータを保護しながら「公開によって価値を生み得るデータ」¹⁾を流通させる仕組みを検討している（新たな情報財検討委員会, 2017）。産学官連携の研究が奨励されていること（田上, 2000）、企業の研究者（企業）によるデータ公開の取り組みも徐々に広がっていること（岡山と岩崎, 2018）から、産学官でデータ公開の課題や解決方法を共有することが望ましいと考えられる。

[4] 分野を横断したデータ公開ポリシーの調査

国内外のオープンサイエンス政策は、分野の状況や特性に配慮しつつ、データ公開を分野横断の課題と位置づけて推進しようとしている（G7 Science and Technology Ministers, 2016; 内閣府, 2017）。しかし、現状では分野を横断してデータの公開状況を把握するマクロな研究はほとんど行われておらず、データベースなどを用いた調査は困難であると考えられる。一方、雑誌のデータ公開ポリシーは、分野ごとのデータ公開の状況がある程度反映していると推測される。雑誌のポリシーは生命科学や政治学など、いくつかの分野を対象として実施されてきたものの（表 2-9）、分野横断の調査は行われてこなかった。

そこで本研究は、雑誌のポリシーの強度をデータ公開の要求状況を反映する指標とみなして22分野を対象とした調査を行った。その結果、ポリシーの掲載率や強度は分野によってばらつきがあること、データ公開が盛んな分野であっても下位の研究領域では差があること、データの再利用（二次分析）が盛んな分野であっても再分析したデータの公開を求める雑誌は多くないことなどがわかった。

[5] データ公開ポリシーと研究不正対策の関連

本研究独自の知見として、雑誌のデータ公開ポリシーには研究不正の抑制が関連している可能性が示唆された。まず、倫理やCOIの開示に関する記述がある雑誌はポリシーの掲載率や強度が高い傾向が示された。間接的に関連があると考えられる特徴として、商業出版社の雑誌はポリシーの掲載率や強度が高い傾向にあった。

¹⁾ 現状ではデータは知的財産権の保護対象ではないが、“収集・蓄積・保管等するために一定の投資又は労力を投じることが必要であり、営業（事業）活動上の利益として保護すべきもの”を「価値あるデータ」と定義している。

また、本研究では雑誌の特徴として、新たに倫理や COI の開示に関する記述、雑誌年齢を調査した。調査の結果、ポリシーの掲載率や強度に有意な差がみられた。ただし、雑誌年齢は低い相関しか認められなかったため、ポリシーを調査する際には雑誌年齢を勘案する必要がないという可能性も示唆された。

[6] データ公開ポリシーの強度と実際のデータ公開状況の比較

Vines et al. (2013b) は、12 誌のデータ公開ポリシーを調査して、要求の強度が高い雑誌ほど、データ公開率が高いことを示した。本研究において、データ公開ポリシーの強度と日本の研究者によるデータの公開状況を分野別に比較した結果、同じく正の相関がみられた。つまり、雑誌のポリシーの強度が高い分野は、データ公開率も高い傾向がみられた。

調査時点では、日本の助成機関や研究機関によるデータ公開要求はほとんどなかったことから、雑誌のデータ公開ポリシーの強度をデータ公開要求の指標とみなして調査した。今後、助成機関や研究機関によるポリシーの策定が進んだ場合は、これらもデータ公開要求の指標として取り入れる必要があると考えられる。

5.3 研究成果の応用

本研究によって、分野ごとのデータ公開の進展状況や課題が明らかになった。学術情報に関する研究や適切な政策立案の一助となるのではないかと考えられる。既に本調査の結果の一部は、オープンサイエンスを推進するための、内閣府や文部科学省学術情報委員会の資料として活用されている。統合イノベーション戦略では、本調査の報告書を先行的な調査として参照した上で、“研究者の公的資金による研究データの公開状況や利活用上の障壁・促進要因等の調査を 2019 年度から実施”するとしている（内閣府, 2018, p. 19）。

また、日本の研究者によるデータ公開の実態や認識を調査したことによって、日本の研究者の課題や特徴を国外の研究者と比較することが可能になった。2018 年に Springer Nature 社が日本の研究者を対象とした質問紙調査を行った際には、本調査の質問項目の一部が採用された（Springer Nature, 2018）。同一の質問項目で調査を行うことによって、調査設計のコストが下がり、より正確な結果の比較も可能になると考えられる。この点は 5.5 で、新たな課題として述べる。

本研究が明らかにした日本におけるデータ公開の主要な課題、や分野別のデータ公開の要求状況とデータ公開の実践状況、データの管理・公開の専門性に関する認識（付表 45 など）は、2.3.3 で述べた図書館等による研究データ管理サービスなど支援体制の検討にも資するのではないだろうか。たとえば、公開用リポジトリやデータ引用など、優先的に支援するべきことからやガイダンスなどに含めることがらが示唆されたと考えられる。

5.4 研究の限界

[1] 質問紙調査

本研究は質問紙調査によって日本の研究者によるデータ公開の実践状況と認識を明らかにした。しかし、データ公開のステークホルダーは研究者だけではなく、政策決定者、助成機関、データ管理者など多様である。Kuipers and van der Hoeven (2009) は、研究者、データ管理者、出版者を対象とした調査を実施して、デジタルデータの保存に関するそれぞれの認識や課題を明らかにしている。ステークホルダー間のギャップを明らかにすることによって、より適切な課題解決の方法を探ることができると考えられる。

データ公開の状況は、既存のデータベースやリポジトリから計量的に調査することが困難であるため、質問紙調査によって明らかにすることとした。できるだけ客観的な公開率に近づけるため、質問紙では「データ公開」の解釈にずれが生じないように、公開方法を選択肢として提示した。しかし、選択肢の意図と回答者の解釈の齟齬は、完全には排除しきれないと推測される。たとえば、データの公開理由では「機関のポリシー」の選択率が予想以上に高かったが、現状では、明文化されたデータ公開ポリシーをもつ研究機関は少ないのではないかと考えられる。各選択肢の表現が適切であったかどうかを改めて検討して、今後の調査に活かしたい。

また、回答者の負担を軽減するために、先行研究の質問や選択肢を一部統合したり、尺度項目を複数選択方式に変更するなどの修正を行った。回答率が高かったことからこれらの修正は有効であったと考えられるが、先行研究と直接比較が行えないという課題も残された。そこで、5.5 に新たな課題として質問項目の標準化について述べる。

調査対象の科学技術専門家ネットワークには企業や公的機関・団体の研究者が含まれており、所属機関による差を明らかにすることができた。一方、回答者の偏り、すなわち工学分野の研究者が多く、宇宙科学や社会科学分野が少ないといった問題が残された。対象者を拡大することによって、より実態に近い分析が可能になると考えられる。

[2] 雑誌のデータ公開ポリシー調査

本研究は、現状ではデータ公開が全ての分野では充分に実施されていないという前提から、各分野で IF が高い雑誌を対象とした。今後、追跡調査をする際には、IF が低～中程度の雑誌を対象とする、あるいは無作為抽出することによって結果を比較したい。

雑誌のデータ公開ポリシーに影響を与える特徴は、分野による差が大きかった。そこで分野グループごとの分析を行ったところ、自然科学は学協会の雑誌の比率や、SP の強度が高いといった傾向がみられた。しかし、分野の影響を完全に制御することは困難であったため、

対象誌数を増やして分野ごとに分析を行うことが考えられる。

また、本研究は研究倫理や COI の開示に関する記述がある雑誌や、商業出版社の雑誌は要求の強度が高いことからデータ公開ポリシーは研究不正の抑制と関連がある可能性が示唆されたと考えた。しかし、これらの値で充分であるとは言い切れないため、新たな指標を検討した上で、データ公開ポリシーとの関連を確認することによって、本研究の結果が妥当であったかどうかを検証したい。

[3] 分野分類

分野の分類には、Essential Science Indicators (ESI) の 22 分類と科学工学指標 (Science and Engineering Indicators, S&EI) に基づく 14 分類を用いた。ESI の雑誌リストは学術雑誌を一意に分類していることや、リストが公開されているため再現や追試が容易であることから採用したが、「人文学」が含まれていない点に限界があると考えられる。データ公開ポリシーの調査時点では、人文学は研究手法や扱うデータの性質が異なることから、自然科学や社会科学分野のデータ公開とは別に議論されることが多かった。しかし、質問紙で S&EI の分類に人文学を追加して調査したところ、データ公開経験をもつ回答者がみられ、他分野からの利用希望もみられた。データ公開ポリシーの追跡調査を行う場合には、人文学、特に学会の創立や議論が盛んになっているデジタルヒューマニティーズ (永崎, 2013) の雑誌も対象として選定したい。

5.5 今後の課題

本研究をふまえた今後の課題として、国際的な質問紙調査項目の開発、継続調査、およびデータ公開ポリシーの効果の検証の 3 点について述べる。

第 1 の課題は、データ公開に関する質問紙調査の標準化である。研究者によるデータ公開の実態や認識に関して、より精密に国際比較を行うことができれば、国や地域ごとの課題が明らかになり、学術情報流通の地域的な特性の解明や、適切な政策の立案に資すると考えられる。本調査では、質問紙調査の項目や分析時の年齢層をできる限り先行研究と揃えることによって国際比較を試みた。しかし、先行研究による調査項目の過不足などによって、項目の一部を変更せざるを得なかった。また、調査項目の作成には労力がかかる。そこで、研究データ同盟 (Research Data Alliance, RDA) の「研究データ共有調査のためのオープン質問」Interest Group (IG) ²⁾で標準化の議論を進めており、ゆくゆくは OECD など国際組織による指

²⁾ Open Questionnaire for Research Data Sharing Survey IG <https://www.rd-alliance.org/groups/open-questionnaire-research-data-sharing-survey-ig>. (accessed 2018-10-31).

標化につなげることを検討している。

第2の課題は、標準化した質問項目を用いて、質問紙調査を継続的に実施することである。2章で述べたように、日本においても2017年にJSTが、2018年にAMEDとNEDOがデータマネジメントプランの提出を要求する方針を公開した。今後、学術コミュニティや社会の変化によって、研究者によるデータ公開の実践や意識が変化すると考えられる。たとえば、データ公開に対する規範の醸成、学術雑誌や助成機関、および研究機関によるデータ公開ポリシーの増加、インフラの開発、データ分析やデータ管理に関する技術の向上、データに関する法の整備、研究不正への対応、市民科学の発展、産学官によるデータの利活用などが、どのような影響を与えるのか、新たな課題は何かといった点を明らかにしていきたい。

第3の課題は、データ公開ポリシーが研究活動にどのような影響をもたらしているかを測定することである。ポリシーによって、データ公開率の向上や撤回論文数の減少といった効果がみられるのかどうかを実証的に明らかにしたい。

学術情報流通におけるデータ公開は揺籃期にあると考えられる。今後、公開データの検索用データベースや引用索引の整備が進めば、計量書誌学などの手法を用いてデータ公開の実践状況を明らかにしたり、分野間の比較を行うことが可能になるであろう。継続的に多角的な方法で研究を行うことによって、データ公開による学術情報流通の変容と社会への影響を捉えていきたい。

謝辞

本研究は、多くの皆様方からのご助力によって遂行することができました。心よりお礼を申し上げます。

まず、本論文の審査をお引き受けくださいました先生方にお礼申し上げます。

指導教員の逸村裕教授（筑波大学）には、博士後期課程への進学から本博士論文の完成まで、常に暖かく親身のご指導を賜りました。研究テーマに関する発表の機会や幅広い方々との交流の機会を設けていただいたことも本研究の基幹となっています。

副指導教員の溝上智恵子教授（筑波大学）、緑川信之名誉教授（筑波大学）には、日頃よりゼミなどを通じて長期にわたりご指導を賜りました。本論文の執筆にあたっては、研究の枠組みから細部に至るまで繰り返し丁寧なご助言を頂戴いたしました。

宇陀則彦准教授（筑波大学）、芳鐘冬樹教授（筑波大学）からは、予備審査を通じて論文の構成や研究課題に関する貴重なご指摘を賜りました。本審査で頂戴したコメントも今後の研究に活かして参りたいと存じます。

竹内比呂也教授（千葉大学）からは、本審査で研究全体に関わるコメントを頂戴しました。日頃よりお伺いしている学術研究に対するご見解、特に論文におけるレビューの重要性とあわせまして、今後の研究の指針とさせていただきます。

また、研究内容について、ご指導やご助言を賜りました先生方にお礼申し上げます。

小野寺夏生名誉教授（筑波大学）には、緑川信之名誉教授との合同研究会にて研究全般、特に統計に関するご助言をいただきました。

杉本重雄特命教授（筑波大学）からは、折に触れてデータをはじめとするデジタル資源の共有、流通について、豊富なご経験に基づく俯瞰的なご助言を賜りました。また、Christine L. Borgman 教授（UCLA）の来日に際しては、雑誌のポリシーについて研究報告の機会をいただき、お二方からご助言を賜ることができました。

辻慶太准教授（筑波大学）には、雑誌のポリシーについての論文執筆の際にご指導を賜りました。論文の構成から順序データの分析まで、多数の貴重なご指摘を頂戴しました。

近藤康久准教授（総合地球環境学研究所）には、オープンチームサイエンスプロジェクトにお誘いいただき、多分野の研究者や政策決定者を交えた研究やワークショップを実施しています。本研究の分野別の分析には、プロジェクトの議論から得られた知見や着想が反映されています。

このほか、多くの先生方から貴重なご助言や示唆をいただいたことが本研究の道標となりました。ここに改めて感謝申し上げます。

謝辞

質問紙調査の実施に際しては、文部科学省科学技術・学術政策研究所（NISTEP）科学技術予測センターの皆様にご助力を賜りました。予備調査および本調査にご協力下さいました皆様にも謹んでお礼申し上げます。

林和弘上席研究官（NISTEP）からは、調査の企画・設計・実施・分析・報告に全面的なご協力を賜りました。また、国内外の政策動向に基づくアドバイスも頂戴しました。

なお、質問紙調査は筑波大学の研究倫理審査の承認を得て実施しました。

筑波大学図書館情報メディア系研究倫理審査（第 28-58 号，2016 年 9 月 6 日）

本研究は、以下の助成を受けて実施しました。ここに謝意を表します。

- (1) 2013 年度日本図書館情報学会研究助成「研究データの分野別公開状況と阻害要因の分析」
- (2) 2013 年度筑波大学・エディンバラ大学研究交流（派遣）「研究データの分野別公開状況と阻害要因：英国におけるデジタルキュレーションの実践」エディンバラ大学／デジタルキュレーションセンター
- (3) 2014 年度三田図書館・情報学会研究助成「高インパクトファクタージャーナルにおける研究データの公開状況と引用への影響」
- (4) 2016 年度総合地球環境学研究所コアプロジェクト FS（14200075）「社会課題解決型研究のアクター間における知識情報ギャップの可視化と克服」

(2)の助成により、本論文の礎となる研究データ公開の動向を英国のデジタルキュレーションセンター（DCC）、エディンバラ大学、グラスゴー大学への訪問調査によって直接確認することができました。インタビューをお引き受け下さいました、Kevin Ashley 氏（DCC）、Angus Whyte 博士（DCC）、Jonathan Rans 氏（DCC）、Stuart Lewis 氏（エディンバラ大学、現・スコットランド国立図書館）、Joy Davidson 氏（グラスゴー大学）に改めてお礼申し上げます。

逸村研究室や小野寺・緑川合同研究会の皆様、溝上研究室の Filiz Yilmaz 氏からは、有益なコメントや励ましをいただきました。筑波大学附属図書館の皆様には、文献複写依頼などで迅速かつ丁寧にご対応いただきました。筑波大学大学院教務の皆様には、博士論文申請に際して親身のご助力をいただきました。深くお礼申し上げます。

お世話になりました全ての方のお名前を挙げることはできませんでしたが、今後の研究で少しずつご恩返していきたいと思います。

最後に、博士課程への進学や論文執筆を応援してくれた家族にお礼を述べさせていただきます。淳さん、舜平太さん、美羽さん、円瀬さん、いつも本当にありがとう。

文献リスト

本論文の引用文献および参考文献を APA 6th edition に準じて示す。ただし，記載方法の一部を以下の通り変更した。

- ・ DOI は直接参照できるように，“<https://doi.org/>”を追記した。
- ・ インターネット上の情報源にはアクセス日付を追記した。
- ・ 電子ジャーナルでページ番号がない文献¹⁾は，PDF 版にページ番号が付されている場合はその番号を記し²⁾，番号にあたるものがない場合は何も記載しなかった³⁾。
- ・ 日本語文献は著者名（フルネーム）を漢字で表記して，欧文著者名の後に五十音順に排列した。本文では“and”を“と”，“et al.”を“ら”と表記し，括弧付き引用では“と”は省略した。

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, 5(4), 594–607. Retrieved from <https://doi.org/10.1016/j.joi.2011.05.007>, (accessed 2018-10-31).

Abrahamson, M., Bollen, K., Gutmann, M. P., King, G., & Pienta, A. (2009). Preserving quantitative research-elicited data for longitudinal analysis: New developments in archiving survey data in the U.S. *Historical Social Research*, 34(3), 51–59. Retrieved from <http://gking.harvard.edu/publications/preserving-data-long-term-analyses>, (accessed 2018-10-31).

Adams, J., & Griliches, Z. (1996). Measuring science: An exploration. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23), 12664–12670. Retrieved from <https://doi.org/10.1073/pnas.93.23.12664>, (accessed 2018-10-31).

Alexandre-Benavent, R., Moreno-Solano, L. M., Sapena, A. F., & Pérez, E. A. (2016). Correlation between impact factor and public availability of published research data in Information Science and Library Science journals. *Scientometrics*, 107(1), 1–13. Retrieved from <https://doi.org/10.1007/s11192-016-1868-7>, (accessed 2018-10-31).

Ali-Khan, S. E., Harris, L. W., & Gold, E. R. (2017). Motivating participation in open science by examining researcher incentives. *eLife*, 6, e29319. Retrieved from

¹⁾ Lee, C. (2015, May 5). How to cite an article with an article number instead of a page range [Web log post]. Retrieved from <https://blog.apastyle.org/apastyle/2015/05/how-to-cite-an-article-with-an-article-number-instead-of-a-page-range.html>, (accessed 2018-12-31).

²⁾ たとえば，Ali-Khan, Harris, and Gold (2017) など。

³⁾ たとえば，Breeze, Poline, and Kennedy (2012) など。

- <https://doi.org/10.7554/eLife.29319>, (accessed 2018-10-31).
- Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. (2011). Public availability of published research data in high-impact journals. *PLOS ONE*, 6(9), e24357. Retrieved from <https://doi.org/10.1371/journal.pone.0024357>, (accessed 2018-10-31).
- Anagnostou, P., Capocasa, M., Milia, N., Sanna, E., Battaglia, C., Luzi, D., & Bisol, G. D. (2015). When data sharing gets close to 100%: What human paleogenetics can teach the open science movement. *PLOS ONE*, 10(3), e0121409. Retrieved from <https://doi.org/10.1371/journal.pone.0121409>, (accessed 2018-10-31).
- Anderson, N. R., Tarczy-Hornoch, P., & Bumgarner, R. E. (2006). On the persistence of supplementary resources in biomedical publications. *BMC Bioinformatics*, 7(1), 260. Retrieved from <https://doi.org/10.1186/1471-2105-7-260>, (accessed 2018-10-31).
- Anderson, R. G., Greene, W. H., McCullough, B. D., & Vinod, H. D. (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology*, 15(1), 99–119. Retrieved from <https://doi.org/10.1080/13501780801915574>, (accessed 2018-10-31).
- André, K. J. (2016). Research data as a global public good. *Journal of Clinical Epidemiology*, 70, 270–271. Retrieved from <https://doi.org/10.1016/j.jclinepi.2015.05.034>, (accessed 2018-10-31).
- Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*, 60(7), 1320–1326. Retrieved from <https://doi.org/10.1002/asi.21062>, (accessed 2018-10-31).
- Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Rebout, L., & Roberge, G. (2013, August). *Proportion of open access peer-reviewed papers at the European and world levels-2004–2011*. Retrieved from http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf, (accessed 2018-10-31).
- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., . . . Wouters, P. (2004). Promoting access to public research data for scientific, economic, and social development. *Data Science Journal*, 3, 135–152. Retrieved from <https://doi.org/10.2481/dsj.3.135>, (accessed 2018-10-31).
- Assante, M., Candela, L., Castelli, D., Manghi, P., & Pagano, P. (2015). Science 2.0 repositories: Time for a change in scholarly communication. *D-Lib Magazine*, 21(1/2). Retrieved from <https://doi.org/10.1045/january2015-assante>, (accessed 2018-10-31).
- Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., . . . Wright, M. H. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon advisory panel on cyberinfrastructure*. Alexandria, VA: National Science Foundation. Retrieved from

- <http://www.nsf.gov/cise/sci/reports/atkins.pdf>, (accessed 2018-10-31).
- Aydinoglu, A. U., Suomela, T., & Malone, J. (2014). Data management in astrobiology: Challenges and opportunities for an interdisciplinary community. *Astrobiology*, 14(6), 451–461. Retrieved from <https://doi.org/10.1089/ast.2013.1127>, (accessed 2018-10-31).
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. Retrieved from <https://doi.org/10.1038/533452a>, (accessed 2018-10-31).
- Bakis, N., Aouad, G., & Kagioglou, M. (2007). Towards distributed product data sharing environments: Progress so far and future challenges. *Automation in Construction*, 16(5), 586–595. Retrieved from <https://doi.org/10.1016/j.autcon.2006.10.002>, (accessed 2018-10-31).
- Bardyn, T. P., Resnick, T., & Camina, S. K. (2012). Translational researchers' perceptions of data management practices and data curation needs: Findings from a focus group in an academic health sciences library. *Journal of Web Librarianship*, 6(4), 274–278. Retrieved from <https://doi.org/10.1080/19322909.2012.730375>, (accessed 2018-10-31).
- Baron, J. N. (1988). Data sharing as a public good. *American Sociological Review*, 53(1), vi–viii.
- Beaver, D. D., & Rosen, R. (1978). Studies in scientific collaboration. *Scientometrics*, 1(1), 65–84. Retrieved from <https://doi.org/10.1007/bf02016840>, (accessed 2018-10-31).
- Berghmans, S., Cousijn, H., Deakin, G., Meijer, I., Mulligan, A., Plume, A., . . . Waltman, L. (2017). *Open data: The researcher perspective*. Centre for Science and Technology Studies (CWTS), Elsevier. Retrieved from <https://www.cwts.nl/download/f-53w2.pdf>, (accessed 2018-10-31).
- Bertagnolli, M. M., Sartor, O., Chabner, B. A., Rothenberg, M. L., Khozin, S., Hugh-Jones, C., . . . Murphy, M. J. (2017). Advantages of a truly open-access data-sharing model. *New England Journal of Medicine*, 376(12), 1178–1181. Retrieved from <https://doi.org/10.1056/NEJMs1702054>, (accessed 2018-10-31).
- Bisol, G. D., Anagnostou, P., Capocasa, M., Bencivelli, S., Cerroni, A., Contreras, J., . . . Boulton, G. (2014). Perspectives on Open Science and scientific data sharing: an interdisciplinary workshop. *Journal of Anthropological Science*, 92, 179–200. Retrieved from <https://doi.org/10.4436/JASS.92006>, (accessed 2018-10-31).
- Björk, B. (2012). The hybrid model for open access publication of scholarly articles: A failed experiment? *Journal of the American Society for Information Science and Technology*, 63(8), 1496–1504. Retrieved from <https://doi.org/10.1002/asi.22709>, (accessed 2018-10-31).
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2010). Sustainable economics for a digital planet: Ensuring long-term access to digital information. Retrieved from <https://www.oclc.org/research/publications/all/long-term-access-digital-information.html>, (accessed 2018-10-31).
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009).

- Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11), 977–984. Retrieved from <https://doi.org/10.1525/bio.2009.59.11.9>, (accessed 2018-10-31).
- Borgman, C. L. (2008). Data, disciplines, and scholarly publishing. *Learned Publishing*, 21(1), 29–38. Retrieved from <https://doi.org/10.1087/095315108x254476>, (accessed 2018-10-31).
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078. Retrieved from <https://doi.org/10.1002/asi.22634>, (accessed 2018-10-31).
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: The MIT Press.
- Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36(1), 2–72. Retrieved from <https://doi.org/10.1002/aris.1440360102>, (accessed 2018-10-31).
- Borgman, C. L., Wallis, J. C., & Mayernik, M. S. (2012). Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work*, 21(6), 485–523. Retrieved from <https://doi.org/10.1007/s10606-012-9169-z>, (accessed 2018-10-31).
- Bornmann, L., & Daniel, H. (2005). Selection of research fellowship recipients by committee peer review: Reliability, fairness and predictive validity of Board of Trustees decisions. *Scientometrics*, 63(2), 297–320. Retrieved from <https://doi.org/10.1007/s11192-005-0214-2>, (accessed 2018-10-31).
- Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80. Retrieved from <https://doi.org/10.1108/00220410810844150>, (accessed 2018-10-31).
- Boulton, G., Rawlins, M., Vallance, P., & Walport, M. (2011). Science as a public enterprise: the case for open data. *The Lancet*, 377(9778), 1633–1635. Retrieved from [https://doi.org/10.1016/S0140-6736\(11\)60647-8](https://doi.org/10.1016/S0140-6736(11)60647-8), (accessed 2018-10-31).
- Bourne, P. E., Clark, T., Dale, R., de Waard, A., Herman, I., Hovy, E., & Shotton, G. (eds.). (2011). *FORCE11 manifesto: Improving future research communication and e-scholarship*. Retrieved from <https://www.force11.org/about/manifesto>, (accessed 2018-10-31).
- Brase, J., Farquhar, A., Gastl, A., Gruttenmeier, H., Heijne, M., Heller, A., . . . Sens, I. (2009). Approach for a joint global registration agency for research data. *Information Services & Use*, 29(1), 13–27. Retrieved from <https://doi.org/10.3233/ISU-2009-0595>, (accessed 2018-10-31).
- Breeze, J. L., Poline, J., & Kennedy, D. N. (2012). Data sharing and publishing in the field of neuroimaging. *GigaScience*, 1(1). Retrieved from <https://doi.org/10.1186/2047-217X-1-9>, (accessed 2018-10-31).
- Brown, R. A., Wolski, M., & Richardson, Joanna. (2015). Developing new skills for research support

- librarians. *Australian Library Journal*, 64(3), 224–234. Retrieved from <https://doi.org/10.1080/00049670.2015.1041215>, (accessed 2018-10-31).
- Bruna, E. M. (2010). Scientific journals can advance tropical biology and conservation by requiring data archiving. *Biotropica*, 42(4), 399–401. Retrieved from <https://doi.org/10.1111/j.1744-7429.2010.00652.x>, (accessed 2018-10-31).
- Callaghan, S. (2014). Preserving the integrity of the scientific record: Data citation and linking. *Learned Publishing*, 27(5), S15–S24. Retrieved from <https://doi.org/10.1087/20140504>, (accessed 2018-10-31).
- Callaghan, S., Tedds, J., Lawrence, R., Murphy, F., Roberts, T., & Wilcox, W. (2014). Cross-linking between journal publications and data repositories: A selection of examples. *International Journal of Digital Curation*, 9(1), 164–175. Retrieved from <https://doi.org/10.2218/ijdc.v9i1.310>, (accessed 2018-10-31).
- Campbell, R., Pentz, E., & Borthwick, I. (2012). *Academic and professional publishing*. Oxford, United Kingdom: Chandos.
- Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747–1762. Retrieved from <https://doi.org/10.1002/asi.23358>, (accessed 2018-10-31).
- Castro, E., Crosas, M., Garnett, A., Sheridan, K., & Altman, M. (2017). Evaluating and promoting open data practices in Open Access journals. *Journal of Scholarly Publishing*, 49(1), 66–88. Retrieved from <https://doi.org/10.3138/jsp.49.1.66>, (accessed 2018-10-31).
- Caulfield, T., Harmon, S. H., & Joly, Y. (2012). Open science versus commercialization: A modern research conflict? *Genome Medicine*, 4(2), 17. Retrieved from <https://doi.org/10.1186/gm316>, (accessed 2018-10-31).
- Center for Open Science (COS). (2014). *Transparency and Openness Promotion (TOP) Guidelines (Version 5)*. Retrieved from https://osf.io/ud578/?_ga=2.267915150.1464883265.1547539130-1295531367.1547539130, (accessed 2018-10-31).
- Center for Research Libraries. (n.d.). *TRAC: The Trustworthy repositories audit & certification: Criteria and checklist*. Retrieved from https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf, (accessed 2018-10-31).
- Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, 19(3), 323–393. Retrieved from https://doi.org/10.1207/s1532690xci1903_3, (accessed 2018-10-31).
- Choi, B. C., Pang, T., Lin, V., Puska, P., Sherman, G., Goddard, M., . . . Clotey, C. (2005). Can scientists and policy makers work together? *Journal of Epidemiology & Community Health*,

- 59(8), 632–637. Retrieved from <https://doi.org/10.1136/jech.2004.031765>, (accessed 2018-10-31).
- Clement, R., Blau, A., Abbaspour, P., & Gandour-Rood, E. (2017). Team-based data management instruction at small liberal arts colleges. *IFLA Journal*, 43(1), 105–118. Retrieved from <https://doi.org/10.1177/0340035216678239>, (accessed 2018-10-31).
- CODATA-ICSTI Task Group on Data Citation Standards and PractOut of Cite, Out of Mind: The Current Sices. (2013). Out of cite, out of mind: the current state of practice, policy, and technology for the citation of data. *Data Science Journal*, 12, CIFCR1–CIFCR75. Retrieved from <https://doi.org/10.2481/dsj.OSOM13-043>, (accessed 2018-10-31).
- Collins, H. M. (1992). *Changing Order: Replication and Induction in Scientific Practice*. Chicago, IL: University of Chicago Press.
- Colquitt, J. A. (2013). From the editors: Data overlap policies at AMJ. *Academy of Management Journal*, 56(2), 331–333. Retrieved from <https://doi.org/10.5465/amj.2013.4002>, (accessed 2018-10-31).
- Consultative Committee for Space Data Systems (CCSDS). (2012). *Recommendation for space data system practices: Reference model for an Open Archival Information System (OAIS)* (CCSDS 650.0-M-2). Retrieved from <https://public.ccsds.org/Pubs/650x0m2.pdf>, (accessed 2018-10-31).
- Conway, E., Giarretta, D., Lambert, S., & Matthews, B. (2011). Curating scientific research data for the long term: A preservation analysis method in context. *International Journal of Digital Curation*, 6(2), 38–52. Retrieved from <https://doi.org/10.2218/ijdc.v6i2.204>, (accessed 2018-10-31).
- Corbyn, Z. (2012). Misconduct is the main cause of life-sciences retractions. *Nature*. 490(7418), 21. Retrieved from <https://doi.org/10.1038/490021a>, (accessed 2018-10-31).
- Costas, R., Meijer, I., Zahedi, Z., & Wouters, P. (2013). *The value of research data: Metrics for datasets from cultural and technical point of view*. Knowledge Exchange, Bristol: United Kingdom. Retrieved from http://repository.jisc.ac.uk/6205/1/Value_of_Research_Data.pdf, (accessed 2018-10-31).
- Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*, 28(8), 454–461. Retrieved from <https://doi.org/10.1016/j.tree.2013.05.002>, (accessed 2018-10-31).
- Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316. Retrieved from <https://doi.org/10.1177/0961000613492542>, (accessed 2018-10-31).
- Cox, A. M., & Verbaan, E. (2016). How academic librarians, IT staff, and research administrators perceive and relate to research. *Library & Information Science Research*, 38(4), 319–326.

- Retrieved from <https://doi.org/10.1016/j.lisr.2016.11.004>, (accessed 2018-10-31).
- Critchlow, T., & Kleese, V. D. (2013). *Data-intensive science*. Boca Raton; London; New York: CRC Press.
- Curry, E., Freitas, A., O'Riain, S. (2010). The role of community-driven data curation for enterprises. In W. David [ed]. *Linking Enterprise Data* (pp. 25–47). Berlin, Germany: Springer.
- Damvad. (2014). *Sharing and Archiving of Publicly Funded Research Data: Report to the Research Council of Norway*. Copenhagen, Norway: Damvad.
- Darby, R., Lambert, S., Matthews, B., Wilson, M., Gitmans, K., Dallmeier-Tiessen, S., . . . Suhonen, J. (2012, October). *Enabling scientific data sharing and re-use*. 2012 IEEE 8th International Conference on E-Science, Chicago, IL. Retrieved from <https://doi.org/10.1109/eScience.2012.6404476>, (accessed 2018-10-31).
- Datas shameful neglect. (2009). *Nature*, 461(7261), 145. Retrieved from <https://doi.org/10.1038/461145a>, (accessed 2018-10-31).
- Davenport, T. H.; Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, October 2012, 70–76. Retrieved from <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>, (accessed 2018-10-31).
- David, P. A. (2003). The economic logic of "open science" and the balance between private property rights and the public domain in scientific data and information: A primer. In National Research Council, *The role of scientific and technical data and information in the public domain: Proceedings of a symposium* (pp. 19–34). Washington, DC: The National Academies Press.
- Davidson, J. (2014). Supporting early-career researchers in data management and curation. In A. Mackenzie & L. Martin (Eds.), *Mastering digital librarianship: Strategy, networking and discovery in academic libraries* (pp. 83–102). London, UNITED KINGDOM: Facet Publishing.
- Day, M. (2008). Toward distributed infrastructures for digital preservation: The roles of collaboration and trust. *International Journal of Digital Curation*, 3(1), 15–28. Retrieved from <https://doi.org/10.2218/ijdc.v3i1.39>, (accessed 2018-10-31).
- Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in empirical economics: The Journal of Money, Credit, and Banking Project. *American Economic Review*, 76(4), 587–603.
- Didier, E., & Guaspare-Cartron, C. (2018). The new watchdogs' vision of science: A roundtable with Ivan Oransky (Retraction Watch) and Brandon Stell (PubPeer). *Social Studies of Science*, 48(1), 165–167. Retrieved from <https://doi.org/10.1177/0306312718756202>, (accessed 2018-10-31).
- Diepenbroek, M., Grobe, H., Reinke, M., Schindler, U., Schlitzer, R., Sieger, R., & Wefer, G. (2002). PANGAEA: An information system for environmental sciences. *Computers & Geosciences*,

- 28(10), 1201–1210. Retrieved from [https://doi.org/10.1016/S0098-3004\(02\)00039-0](https://doi.org/10.1016/S0098-3004(02)00039-0), (accessed 2018-10-31).
- Digital Curation Centre (DCC). (n. d.). Overview of funders' data policies. Retrieved from <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>, (accessed 2018-10-31).
- Doorn, P., & Tjalsma, H. (2007). Introduction: Archiving research data. *Archival Science*, 7(1), 1–20. Retrieved from <https://doi.org/10.1007/s10502-007-9054-6>, (accessed 2018-10-31).
- Dorch, B. F., Drachen, T. M., & Ellegaard, O. (2015). The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union*, 11(A29A), 172–175. Retrieved from <https://doi.org/10.1017/S1743921316002696>, (accessed 2018-10-31).
- Dörnyei, Z. (2006). *Questionnaires in second language research: Construction, administration, and processing* 外国語教育学のための質問紙調査入門: 作成・実施・データ処理 (八島智子, 竹内理, Trans.). 東京, 日本: 松柏社. (Original work published 2003).
- Doshi, P., Goodman, S. N., Ioannidis, J. P. (2013). Raw data from clinical trials: Within reach? *Trends in Pharmacological Sciences*, 34(12), 645–647. Retrieved from <https://doi.org/10.1016/j.tips.2013.10.006>, (accessed 2018-10-31).
- Dryad. (2011). Joint Data Archiving Policy (JDAP). Retrieved from <http://datadryad.org/pages/jdap>, (accessed 2018-10-31).
- Dryad. (2014). *2014 Organizational and membership report*. Retrieved from <http://datadryad.org/themes/Mirage/docs/DryadAnnualReport2014.pdf>, (accessed 2018-10-31).
- Edwards, A. (2016). Science is still too closed. *Nature*, 533, s70. Retrieved from <https://doi.org/10.1038/533S70a>, (accessed 2018-10-31).
- Edwards, P. N., Jackson, S. J., Bowkerm G. C., & Knobel, C. P. (2007). *Understanding infrastructure: Dynamics, tensions, and design. Report of a workshop on “History & theory of infrastructure: Lessons for new scientific cyberinfrastructures”*. Retrieved from <http://hdl.handle.net/2027.42/49353>, (accessed 2018-10-31).
- Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: data, metadata, and collaboration. *Social Studies of Science*, 41(5), 667–690. Retrieved from <https://doi.org/10.1177/0306312711413314>, (accessed 2018-10-31).
- Elsevier. (2016). Data Citation. Retrieved from <https://www.elsevier.com/about/open-science/research-data/data-citation>, (accessed 2018-10-31).
- Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B., & Gemeinholzer, B. (2012). The user's view on biodiversity data sharing: Investigating facts of acceptance and requirements to realize a sustainable use of research data. *Ecological Informatics*, 11, 25–33. Retrieved from <https://doi.org/10.1016/j.ecoinf.2012.03.004>, (accessed 2018-10-31).

- Environmental Protection Agency (EPA). (2015, March 13). Memorandum of understanding information sharing between the United States Environmental Protection Agency Office of Chemical Safety and Pollution Prevention and Department of Health and Human Services Food and Drug Administration Food and Veterinary Medicine Program. Retrieved from <http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2003-0004-0095>, (accessed 2018-10-31).
- European Commission (EC). (2013). *Guidelines on Data Management in Horizon 2020 (Ver. 1.0)*. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, (accessed 2018-10-31).
- European Commission (EC). (2015). *Validation of the results of the public consultation on Science 2.0: Science in transition*. Retrieved from <https://ec.europa.eu/digital-single-market/news/final-report-science-20-public-consultation>, (accessed 2018-10-31).
- European Commission (EC). (2016). *E-infrastructures: Making Europe the best place for research and innovation*. Retrieved from <https://ec.europa.eu/digital-single-market/en/news/e-infrastructures-making-europe-best-place-research-and-innovation>, (accessed 2018-10-31).
- European Commission (EC). (2017, October 26). *EOSC Declaration*. Retrieved from https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf, (accessed 2018-10-31).
- European Commission (EC). (2018). Open Science Monitor. Retrieved from <http://ec.europa.eu/research/openscience/index.cfm?pg=home§ion=monitor>, (accessed 2018-10-31).
- European Medicines Agency (EMA). (2014, October 2). *European Medicines Agency Policy on Publication of Clinical Data for Medicinal Products for Human Use*. [Effective date: 2015-1-1]. Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf, (accessed 2018-10-31).
- European Union. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data: Final report of the High Level Expert Group on scientific data; A submission to the European Commission*. Retrieved from <https://www.fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data>, (accessed 2018-10-31).
- Evangelou, E., Trikalinos, T. A., & Ioannidis, J. P. (2005). Unavailability of online supplementary scientific information from articles published in major journals. *The FASEB Journal*, 19(14), 1943–1944. Retrieved from <https://doi.org/10.1096/fj.05-4784lsf>, (accessed 2018-10-31).
- Expert Advisory Group on Data Access, Economic and Social Research Council, Medical Research Council, Cancer Research UK, & Wellcome Trust. (2014). *Establishing incentives and changing cultures to support data access*. Retrieved from

- <http://wellcomelibrary.org/item/b20928129#>, (accessed 2018-10-31).
- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42), 17208–17033. Retrieved from <https://doi.org/10.1073/pnas.1212247109>, (accessed 2018-10-31).
- Faniel, I. M., Kriesberg, A., & Yakel, E. (2016). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, 67(6), 1404–1416. Retrieved from <https://doi.org/10.1002/asi.23480>, (accessed 2018-10-31).
- Fearon, D. Jr., Gunia, B., Lake, S., Pralle, B. E., Sallans, A. L. (2013). *Research Data Management services*(SPEC Kit 334). Washington, DC: Association of Research Libraries (ARL). Retrieved from <https://doi.org/10.29242/spec.334>, (accessed 2018-10-31).
- Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLOS ONE*, 10(2), e0118053. Retrieved from <https://doi.org/10.1371/journal.pone.0118053>, (accessed 2018-10-31).
- Federer, L. (2018). Defining data librarianship: A survey of competencies, skills, and training. *Journal of the Medical Library Association*, 106(3), 294–303. Retrieved from <https://doi.org/10.5195/JMLA.2018.306>, (accessed 2018-10-31).
- Feigelson, E. D., & Babu, G. J. (2012). Big data in astronomy. *Significance*, 9(4), 22–25. Retrieved from <https://doi.org/10.1111/j.1740-9713.2012.00587.x>, (accessed 2018-10-31).
- Ferguson, L. (2014, November 3). How and why researchers share data (and why they don't) [Web log post]. Retrieved from <https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont>, (accessed 2018-10-31).
- Field, D., Sansone, S., Collis, A., Booth, T., Dukes, P., Gregurick, S. K., . . . Wilbanks, J. (2009). 'Omics data sharing. *Science*, 326(5950), 234–236. Retrieved from <https://doi.org/10.1126/science.1180598>, (accessed 2018-10-31).
- Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). *Sharing Research Data*. Washington, DC: National Academies Press.
- FORCE11. (2014). Joint Declaration of Data Citation Principles (JDDCP). Retrieved from <https://www.force11.org/group/joint-declaration-data-citation-principles-final>, (accessed 2018-10-31).
- FORCE11. (2016). The FAIR Data Principles. Retrieved from <https://www.force11.org/group/fairgroup/fairprinciples>, (accessed 2018-10-31).
- G7 Science and Technology Ministers. (2016). *G7 Tsukuba Communiqué: つくばコミュニケ (共同声明)*. Retrieved from http://www8.cao.go.jp/cstp/kokusaiteki/g7_2016/2016communique.html, (accessed 2018-10-

- 31).
- G7 Science Ministers. (2017). *G7 Science Ministers' Communiqué*. Retrieved from http://www8.cao.go.jp/cstp/kokusaiteki/g7_2017/20170928communiqué.pdf, (accessed 2018-10-31).
- G8 Science Ministers. (2013, June 13). *G8 Science Ministers Statement*. Retrieved from <https://www.gov.uk/government/news/g8-science-ministers-statement>, (accessed 2018-10-31).
- Gault, B. & Koers, H. (2015, February 4). Supporting data-sharing to speed up innovation in Materials Science: Elsevier's open data initiative will provide new ways of storing, sharing and accessing research data [Web log post]. Retrieved from <http://www.elsevier.com/connect/supporting-data-sharing-to-speed-up-innovation-in-materials-science>, (accessed 2018-10-31).
- Gherghina, S., & Katsanidou, A. (2013). Data availability in political science journals. *European Political Science*, 12(3), 333–349. Retrieved from <https://doi.org/10.1057/eps.2013.8>, (accessed 2018-10-31).
- Gleditsch, N. P., Metelits, C., & Strand, H. (2003). Posting your data: Will you be scooped, or will you be famous? *International Studies Perspectives*, 4(1), 89–97.
- Greenberg, S. A. (2009). How citation distortions create unfounded authority: Analysis of a citation network. *BMJ*, 339(Jul20 3), b2680. Retrieved from <https://doi.org/10.1136/bmj.b2680>, (accessed 2018-10-31).
- Guttmacher, A. E., Nabel, E. G., Collins, F. S. (2009). Why data-sharing policies matter. *Proceedings of the National Academy of Sciences of the United States of America*, 106(40), 16894. Retrieved from <https://doi.org/10.1073/pnas.0910378106>, (accessed 2018-10-31).
- Hachani, S. (2015). Open peer review: Fast forward for a new science. *Current issues in libraries, information science and related fields advances in librarianship*, 39, 115-141. Retrieved from <https://doi.org/10.1108/s0065-283020150000039012>, (accessed 2018-10-31).
- Harzing, A., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. Retrieved from <https://doi.org/10.1007/s11192-015-1798-9>, (accessed 2018-10-31).
- Heckerman, D., Takeda, K., Emmott, S., Hey, T. & Xu, Y. (eds.). (2012). *Science@Microsoft: The fourth paradigm in practice*. Redmond, WA: Microsoft corporation.
- Higher Education Funding Council (HEFCE), Research Council UK (RCUK), Universities UK (UUK), & Wellcome Trust. (2016, July 28). *Concordat on open research data*. Retrieved from <https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/>, (accessed 2018-10-31).
- Henneken, E. (2015). Unlocking and sharing data in astronomy. *Bulletin of the Association for*

- Information Science and Technology*, 41(4), 40–43. Retrieved from <https://doi.org/10.1002/bult.2015.1720410412>, (accessed 2018-10-31).
- Henneken, E. A., & Accomazzi, A. (2012). Linking to data: Effect on citation rates in astronomy. *Astronomical Data Analysis Software and Systems XXI, ASP Concerence Series*, 461, 763–766. Retrieved from <http://www.aspbooks.org/publications/461/763.pdf>, (accessed 2018-10-31).
- Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery* (1st ed.). Redmond, WA: Microsoft Research. Retrieved from <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>, (accessed 2018-10-31).
- Hey, T., & Trefethen, A. (2003). The data deluge: An e-Science perspective. In F. Berman, G. Fox, & T. Hey. (Eds.), *Grid computing: Making the global infrastructure a reality* (pp. 809–824). Chichester, United Kingdom: Wiley.
- Horton, L., & Katsanidou, A. (2011). Purposing your survey: Archives as a market regulator, or how can archives connect supply and demand? *IASSIST Quarterly*, 35(4), 18–23. Retrieved from http://www.iassistdata.org/sites/default/files/iqvol35_horton.pdf, (accessed 2018-10-31).
- Hottenrott, H., & Lawson, C. (2017). A first look at multiple institutional affiliations: A study of authors in Germany, Japan and the UK. *Scientometrics*, 111(1), 285–295. Retrieved from <https://doi.org/10.1007/s11192-017-2257-6>, (accessed 2018-10-31).
- Houghton, J., & Gruen, N. (2014). *Open research data: Report to the Australian National Data Service (ANDS)*. Retrieved from https://www.ands.org.au/__data/assets/pdf_file/0019/393022/open-research-data-report.pdf, (accessed 2018-10-31).
- Howard, A. (2012). *Data for the public good: How data can help citizens and government*. Sebastopol, CA: O'reilly.
- Howlett, M. (2009). Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration*, 52(2), 153–175. Retrieved from <https://doi.org/10.1111/j.1754-7121.2009.00070.x>, (accessed 2018-10-31).
- Huang, X., Hawkins, B. A., Lei, F., Miller, G. L., Favret, C., Zhang, R., & Qiao, G. (2012). Willing or unwilling to share primary biodiversity data: Results and implications of an international survey. *Conservation Letters*, 5(5), 399–406. Retrieved from <https://doi.org/10.1111/j.1755-263X.2012.00259.x>, (accessed 2018-10-31).
- Hyman, H. H. (1972). *Secondary analysis of sample surveys: Principles, procedures, and potentialities*. New York, NY: Wiley.
- Ikeuchi, U. (2016). *Data sharing policies in scholarly journals across 22 disciplines* [Dataset]. Retrieved from <https://doi.org/10.6084/m9.figshare.3144991>, (accessed 2018-10-31).
- Ikeuchi, U., & Itsumura, H. (2015). A snapshot of data citation using the Data Citation Index and the

- Web of Science. 10th International Digital Curation Conference (IDCC 2015). London, February 9-12, 2015.
- InterPARES 2 Project. (2008). *General study 10 final report: Preservation practices of scientific data portals*. Retrieved from http://www.interpares.org/display_file.cfm?doc=ip2_gs10_final_report.pdf, (accessed 2018-10-31).
- Ioannidis, J. P. (2014). How to make more published research true. *PLOS Medicine*, 11(10), e1001747. Retrieved from <https://doi.org/10.1371/journal.pmed.1001747>, (accessed 2018-10-31).
- Kane, E. J. (1984). Why journal editors should encourage the replication of applied econometric research. *Quarterly Journal of Business & Economics*, 23 (1), 3–8.
- Kaye, J., Heeney, C., Hawkins, N., Vries, J. D., & Boddington, P. (2009). Data sharing in genomics: Reshaping scientific practice. *Nature Reviews Genetics*, 10, 331–335. Retrieved from <https://doi.org/10.1038/nrg2573>, (accessed 2018-10-31).
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L., . . . Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, 14(5). Retrieved from <https://doi.org/10.1371/journal.pbio.1002456>, (accessed 2018-10-31).
- King, D. W., McDonald, D. D., & Roderer, N. K. (1981). *Scientific journals in the United States: Their production, use, and economics*. Stroudsburg, PA: Hutchinson Ross Pub.
- Kuipers, T. & van der Hoeven, J. (2009). *Insight into digital preservation of research output in Europe*. PARSE.Insight (INSIGHT into issues of Permanent Access to the Records of Science in Europe). Retrieved from <https://libereurope.eu/wp-content/uploads/2010/01/PARSE.Insight-Deliverable-D3.4-Survey-Report.-of-research-output-Europe-Title-of-Deliverable-Survey-Report.pdf>, (accessed 2018-10-31).
- Jones, M. B., Schildhauer, M. P., Reichman, O. J., & Bowers, S. (2006). The new bioinformatics: Integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, 37, 519–544. Retrieved from <https://doi.org/10.1146/annurev.ecolsys.37.091305.110031>, (accessed 2018-10-31).
- Jones, S. (2011). *How to develop a data management and sharing plan*. Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf>, (accessed 2018-10-31).
- Jones, S., Pryor, G., & Whyte, A. (2013). *How to develop research data management services: A guide for HEIs*. Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/publications/How-to-develop-RDM-services_finalMay2013rev.pdf, (accessed 2018-10-31).

- Jubb, M. (2013). Introduction: Scholarly communications - Disruptions in a complex ecology. In D. Shorley & M. Jubb (Eds.), *The Future of Scholarly Communication* (pp. XII–XXXVI). London, United Kingdom: Facet Publishing.
- Kafkas, Ş, Kim, J., Pi, X., & Mcentyre, J. R. (2015). Database citation in supplementary data linked to Europe PubMed Central full text biomedical articles. *Journal of Biomedical Semantics*, 6(1), 1. Retrieved from <https://doi.org/10.1186/2041-1480-6-1>, (accessed 2018-10-31).
- Kahn, M., Higgs, R., Davidson, J., & Jones, S. (2014). Research data management in South Africa: How we shape up. *Australian Academic & Research Libraries*, 45(4), 296–308. Retrieved from <https://doi.org/10.1080/00048623.2014.951910>, (accessed 2018-10-31).
- Kane, E. J. (1984). Why journal editors should encourage the replication of applied econometric research. *Quarterly Journal of Business & Economics*, 23(1), 3–8.
- Karasti, H., Baker, K. S., & Halkola, E. (2006). Enriching the notion of data curation in e-science: Data managing and information infrastructuring in the Long Term Ecological Research (LTER) network. *Computer Supported Cooperative Work (CSCW)*, 15(4), 321–358.
- Kelty, C. M. (2008). *Two bits: The cultural significance of free software*. Durham, NC: Duke University Press.
- Kenyon, J., Sprague, N., & Flathers, E. (2016). The journal article as a means to share data: A content analysis of supplementary materials from two disciplines. *Journal of Librarianship and Scholarly Communication*, 4, eP2112. Retrieved from <https://doi.org/10.7710/2162-3309.2112>, (accessed 2018-10-31).
- Kim, Y., & Burns, C. S. (2015). Norms of data sharing in biological sciences: The roles of metadata, data repository, and journal and funding requirements. *Journal of Information Science*, 42(2), 230–245. Retrieved from <https://doi.org/10.1177/0165551515592098>, (accessed 2018-10-31).
- Kim, Y., & Kim, S. (2015). Institutional, motivational, and resource factors influencing health scientists' data-sharing behaviors. *Journal of Scholarly Publishing*, 46(4), 366–389. Retrieved from <https://doi.org/10.3138/jsp.46.4.05>, (accessed 2018-10-31).
- Kim, Y., & Stanton, J. M. (2015). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, 67(4), 776–799. Retrieved from <https://doi.org/10.1002/asi.23424>, (accessed 2018-10-31).
- Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*, 37(3), 189–200. Retrieved from <https://doi.org/10.1016/j.lisr.2015.04.006>, (accessed 2018-10-31).
- King, G. (1995). Replication, replication. *Political Science and Politics*, 28(3), 444–452. Retrieved from <http://www.jstor.org/stable/420301>, (accessed 2018-10-31).
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., . . . Nosek, B.

- A. (2014). Investigating variation in replicability: A 'Many Labs' replication project. *Social Psychology*, 45(3), 142–152. Retrieved from <https://doi.org/10.1027/1864-9335/a000178>, (accessed 2018-10-31).
- Knowledge Exchange. (2017a). *Knowledge Exchange approach to open scholarship*. Retrieved from <https://doi.org/10.5281/zenodo.826643>, (accessed 2018-10-31).
- Knowledge Exchange. (2017b). *The evolving landscape of federated research data infrastructures*. Retrieved from <https://doi.org/10.5281/zenodo.1064730>, (accessed 2018-10-31).
- Kowalczyk, S., & Shankar, K. (2013). Data sharing in the sciences. *Annual Review of Information Science and Technology*, 45(1), 247–294. Retrieved from <https://doi.org/10.1002/aris.2011.1440450113>, (accessed 2018-10-31).
- Kratz, J. E., & Strasser, C. (2015a). Researcher perspectives on publication and peer review of data. *PLOS ONE*, 10(2), e0117619. Retrieved from <https://doi.org/10.1371/journal.pone.0117619>, (accessed 2018-10-31).
- Kratz, J. E., & Strasser, C. (2015b). Making data count. *Scientific Data*, 2, 150039 (2015). Retrieved from <https://doi.org/10.1038/sdata.2015.39>, (accessed 2018-10-31).
- Lafferty-Hess, S., Rudder, J., Downey, M., Ivey, S., & Darragh, J. (2018, May 30). Conceptualizing data curation activities within two academic libraries. [LIS Scholarship Archive Preprints]. Retrieved from <https://doi.org/10.17605/OSF.IO/ZJ5PQ>, (accessed 2018-10-31).
- Lakomaa, E., & Kallberg, J. (2013). Open data as a foundation for innovation: The enabling effect of free public sector information for entrepreneurs. *IEEE Access*, 1, 558–563. Retrieved from <https://doi.org/10.1109/ACCESS.2013.2279164>, (accessed 2018-10-3).
- Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and peer review of data: Moving towards formal data publication. *The International Journal of Digital Curation*, 6(2), 437. Retrieved from <https://doi.org/10.2218/ijdc.v6i2.205>, (accessed 2018-10-31).
- Lehman, R., & Loder, E. (2012). Missing clinical trial data: A threat to the integrity of evidence based medicine. *BMJ*, 344. Retrieved from <https://doi.org/10.1136/bmj.d8158>, (accessed 2018-10-31).
- League of European Research Universities (LERU). (2012). *LERU statement on Open Research Data*. Retrieved from <https://www.leru.org/files/LERU-Statement-on-Open-Research-Data-Full-paper.pdf>, (accessed 2018-10-31).
- Leydesdorff, L., & Bornmann, L. (2015). The operationalization of “fields” as WoS subject categories (WCs) in evaluative bibliometrics: The cases of “library and information science” and “science & technology studies”. *Journal of the Association for Information Science and Technology*, 67(3), 707–714. Retrieved from <https://doi.org/10.1002/asi.23408>, (accessed 2018-10-31).
- Lindsay, D. S. (2017). Sharing data and materials in psychological science. *Psychological Science*,

- 28(6), 699–702. Retrieved from <https://doi.org/10.1177/0956797617704015>, (accessed 2018-10-31).
- Liu, Y., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology*, 61(2), 340–351. Retrieved from <https://doi.org/10.1002/asi.21248>, (accessed 2018-10-31).
- Luukkonen, T., Tijssen, R. J., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15–36. Retrieved from <https://doi.org/10.1007/bf02016282>, (accessed 2018-10-31).
- Malin, B., Karp, D., & Scheuermann, R. H. (2010). Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. *Journal of Investigative Medicine*, 58(1), 11–18. Retrieved from <https://doi.org/10.2310/JIM.0b013e3181c9b2ea>, (accessed 2018-10-31).
- Martinez, M. G., Verbruggen, P., & Fearn, A. (2013). Risk-based approaches to food safety regulation: What role for co-regulation? *Journal of Risk Research*, 16(9), 1101–1121. Retrieved from <https://doi.org/10.1080/13669877.2012.743157>, (accessed 2018-10-31).
- Martone, M. E. (2013). Brain and Behavior: We want you to share your data. *Brain and Behavior*, 4(1), 1–3. Retrieved from <https://doi.org/10.1002/brb3.192>, (accessed 2018-10-31).
- Maunsell, J. (2010). Announcement regarding supplemental material. *The Journal of Neuroscience*, 30(32), 10599–10600. Retrieved from <http://www.jneurosci.org/content/30/32/10599.full.pdf+html>, (accessed 2018-10-31).
- Mauthner, N. S., & Parry, O. (2013). Open access digital data sharing: Principles, policies and practices. *Social Epistemology*, 27(1), 47–67. Retrieved from <https://doi.org/10.1080/02691728.2012.760663>, (accessed 2018-10-30).
- Maxmen, A. (2018). AI researchers embrace Bitcoin technology to share medical data. *Nature*, 555, 293–294. Retrieved from <https://doi.org/10.1038/d41586-018-02641-7>, (accessed 2018-10-31).
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, 70(6), 487–498. Retrieved from <https://doi.org/10.1037/a0039400>, (accessed 2018-10-31).
- Mayo, C., Vision, T. J., & Hull, E. A. (2016). The location of the citation: Changing practices in how publications cite original data in the Dryad Digital Repository. *International Journal of Digital Curation*, 11(1), 150. Retrieved from <https://doi.org/10.2218/ijdc.v11i1.400>, (accessed 2018-10-31).
- McCain, K. W. (1995). Mandating sharing: Journal policies in the natural sciences. *Science Communication*, 16(4), 403–431.

- McCain, K. W. (2000). Sharing digitized research-related information on the World Wide Web. *Journal of the American Society for Information Science*, 51(14), 1321–1327. Retrieved from [https://doi.org/10.1002/1097-4571\(2000\)9999:9999::AID-ASII048>3.0.CO;2-Q](https://doi.org/10.1002/1097-4571(2000)9999:9999::AID-ASII048>3.0.CO;2-Q), (accessed 2018-10-31).
- McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2008). Do economics journal archives promote replicable research? *Canadian Journal of Economics*, 41(4), 1406–1420.
- Mckiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., . . . Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, 5, e16800. Retrieved from <https://doi.org/10.7554/eLife.16800>, (accessed 2018-10-31).
- Merriam, S. B. (2004). *Qualitative research and case study applications in education 質的調査法入門：教育における調査法とケーススタディ*. 叢書・現代社会のフロンティア, 3. (堀薫夫, 久保真人, 成島美弥, Trans.). 東京, 日本: ミネルヴァ書房. (Original work published 1998).
- Midorikawa, N., Ogawa, H., Saito, K., Kaneko, M., & Itsumura, H. (1984). The relationship among the citation measures and the factors influence on them. *Information Services & Use*, 4, 417–424.
- Moles, N. (2015). Data-PE: A framework for evaluating data publication policies at scholarly journals. *Data Science Journal*, 13, 192–202. Retrieved from <https://doi.org/10.2481/dsj.14-047>, (accessed 2018-10-31).
- Molloy, J. C. (2011). The Open Knowledge Foundation: Open Data Means Better Science. *PLOS Biology*, 9(12), e1001195. Retrieved from <https://doi.org/10.1371/journal.pbio.1001195>, (accessed 2018-10-31).
- Monastersky, R. (2013). Publishing frontiers: The library reboot. *Nature*, 495(7442), 430–432. Retrieved from <https://doi.org/10.1038/495430a>, (accessed 2018-10-31).
- Montero-Melendez, T., & Perretti, M. (2014). Connections in pharmacology: innovation serving translational medicine. *Drug Discovery Today*, 19(7), 820–823. Retrieved from <https://doi.org/10.1016/j.drudis.2013.11.022>, (accessed 2018-10-31).
- Murray-Rust, P., Neylon, C., Pollock, R., & Wilbanks, J. (2010, February 19). *Panton principles: Principles for open data in science*. Retrieved from <https://pantonprinciples.org>, (accessed 2018-10-31).
- National Institutes of Health (NIH). (2003, February 26). *Final NIH statement on sharing research data*. Retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>, (accessed 2018-10-31).
- National Institutes of Health (NIH). (2004). Frequently Asked Questions: Data sharing. Retrieved from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm, (accessed 2018-10-31).

- National Information Standards Organization (NISO) and the National Federation of Advanced Information Services (NFAIS). (2013). *Recommended practices for online supplemental journal article materials*. Retrieved from http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf, (accessed 2018-10-31).
- National Research Council (NRC). (1997). *Bits of power: Issues in global access to scientific data*. Washington, DC: The National Academies Press. Retrieved from <https://doi.org/10.17226/5504>, (accessed 2018-10-31).
- National Research Council (NRC), & Global Affairs. (2002, April 22). *Access to research data in the 21st century: An ongoing dialogue among interested parties: Report of a workshop*. Washington, DC: National Academies Press. Retrieved from <https://www.nap.edu/catalog/10302/access-to-research-data-in-the-21st-century-an-ongoing>, (accessed 2018-10-31).
- National Science Board (NSB). (2001). The Board's bestseller. In *The National Science Board: A History in Highlights 1950-2000* (p. 20). Retrieved from <https://nsf.gov/nsb/documents/2000/nsb00215/nsb00215.pdf>, (accessed 2018-10-31).
- National Science Board (NSB). (2005). Long-lived digital data collections: Enabling research and education in the 21st century. Arlington, VA: National Science Foundation. Retrieved from <https://www.nsf.gov/pubs/2005/nsb0540/>, (accessed 2018-10-31).
- National Science Board (NSB). (2018). Science & Engineering Indicators 2018. Retrieved from <https://www.nsf.gov/statistics/2018/nsb20181/>, (accessed 2018-10-31).
- National Science Foundation (NSF). (2011). Chapter II – Proposal preparation instructions. *Proposal & award policies & procedures guide*. Retrieved from https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_2.jsp#IIC2j, (accessed 2018-10-31).
- National Science Foundation (NSF). (2013, January). GPG summary of changes. Retrieved from https://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_sigchanges.jsp, (accessed 2018-10-31).
- Naudet, F., Sakarovitch, C., Janiaud, P., Cristea, I., Fanelli, D., Moher, D., & Ioannidis, J. P. (2018). Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: Survey of studies published in The BMJ and PLOS Medicine. *BMJ*, 360, k400. Retrieved from <https://doi.org/10.1136/bmj.k400>, (accessed 2018-10-31).
- Nelson, B. (2009). Data sharing: Empty archives. *Nature*, 461, 160–163. Retrieved from <https://doi.org/10.1038/461160a>, (accessed 2018-10-31).
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. Retrieved from <https://doi.org/10.1126/science.aab2374>, (accessed 2018-10-31).

- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631. Retrieved from <https://doi.org/10.1177/1745691612459058>, (accessed 2018-10-31).
- Open Knowledge International. (n.d.). The Open Definition. Retrieved from <http://opendefinition.org>, (accessed 2018-10-31).
- Open Research Data Task Force. (2017). *Research data infrastructures in the UK*. London, United Kingdom: Universities UK.
- Organisation for Economic Co-operation and Development (OECD). (2004, January 30). *Declaration on access to research data from public funding*. Retrieved from <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157>, (accessed 2018-10-31).
- Organisation for Economic Co-operation and Development (OECD). (2007). *OECD principles and guidelines for access to research data from public funding*. Paris, France: OECD Publications. Retrieved from <http://www.oecd.org/science/sci-tech/38500813.pdf>, (accessed 2018-10-31).
- Organisation for Economic Co-operation and Development (OECD). (2012). “Gross Domestic Expenditure on R&D: GERD (million current PPP \$)”, “GERD as a percentage of GDP”. In *Main Science and Technology Indicators. StatExtracts*. Retrieved from http://stats.oecd.org/Index.aspx?DataSetCode=MSTI_PUB, (accessed 2018-10-31).
- Organisation for Economic Co-operation and Development (OECD). (2014). *Measuring the digital economy: A new perspective*. Paris, France: OECD Publishing. Retrieved from <https://doi.org/10.1787/9789264221796-en>, (accessed 2018-10-31).
- Organisation for Economic Co-operation and Development (OECD). (2015). *Making open science a reality* (OECD Science, Technology and Industry Policy Papers, No. 25). Paris, France: OECD Publishing. Retrieved from <https://doi.org/10.1787/5jrs2f963zs1-en>, (accessed 2018-10-31).
- Organisation for Economic Co-operation and Development (OECD). (2017). *Business models for sustainable research data repositories* (OECD Science, Technology and Industry Policy Papers, No. 47). Paris, France: OECD Publishing. Retrieved from <https://doi.org/10.1787/302b12bb-en>, (accessed 2018-10-31).
- Office of Science and Technology Policy (OSTP). (2013, February 22). *Increasing access to the results of federally funded scientific research*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf, (accessed 2018-10-31).
- Office of Science and Technology Policy (OSTP). (2014, June 1). *Open government plan*. Retrieved from <https://obamawhitehouse.archives.gov/open/around/eop/ostp/plan>, (accessed 2018-10-31).

- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., . . . Dierolf, U. (2013). Making research data repositories visible: The re3data.org registry. *PLOS ONE*, 8(11), e78080. Retrieved from <https://doi.org/10.1371/journal.pone.0078080>, (accessed 2018-10-31).
- Parsons, M. A., Duerr, R., & Minster, J. (2010). Data citation and peer review. *Eos, Transactions American Geophysical Union*, 91(34), 297–298. Retrieved from <https://doi.org/10.1029/2010EO340001>, (accessed 2018-10-31).
- Pasquetto, I. (2018). *From open data to knowledge production: Biomedical data sharing and unpredictable data reuses* (Doctoral Dissertation, University of California, Los Angeles). Retrieved from <https://escholarship.org/uc/item/1s1814cj>, (accessed 2018-12-11).
- Peer, L., Green, A., & Stephenson, E. (2014). Committing to data quality review. *International Journal of Digital Curation*, 9(1), 263–291. Retrieved from <https://doi.org/10.2218/ijdc.v9i1.317>, (accessed 2018-10-31).
- Pepe, A., Goodman, A., Muench, A., Crosas, M., & Erdmann, C. (2014). How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers. *PLOS ONE*, 9(8), e104798. Retrieved from <https://doi.org/10.1371/journal.pone.0104798>, (accessed 2018-10-31).
- Perrino, T., Howe, G., Sperling, A., Beardslee, W., Sandler, I., Shern, D., . . . Brown, H. C. (2013). Advancing science through collaborative data sharing and synthesis. *Perspectives on Psychological Science*, 8(4), 433–444. Retrieved from <https://doi.org/10.1177/1745691613491579>, (accessed 2018-10-31).
- Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. (2016). Research data explored: An extended analysis of citations and altmetrics. *Scientometrics*, 107(2), 723–744. Retrieved from <https://doi.org/10.1007/s11192-016-1887-4>, (accessed 2018-10-31).
- Pham-Kanter, G., Zinner, D. E., & Campbell, E. G. (2014). Codifying collegiality: Recent developments in data sharing policy in the life sciences. *PLOS ONE*, 9(9), e108451. Retrieved from <https://doi.org/10.1371/journal.pone.0108451>, (accessed 2018-10-31).
- Phillips, N. (2017). Striving for a research renaissance. *Nature*, 543(7646), S7 (Nature Index). Retrieved from <https://doi.org/10.1038/543S7a>, (accessed 2018-10-31).
- Perrier, L., Blondal, E., Ayala, A. P., Dearborn, D., Kenny, T., Lightfoot, D., . . . Macdonald, H. (2017). Research data management in academic institutions: A scoping review. *Plos One*, 12(5). Retrieved from <https://doi.org/10.1371/journal.pone.0178261>, (accessed 2018-10-31).
- Piowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLOS ONE*, 6(7), e18657. Retrieved from <https://doi.org/10.1371/journal.pone.0018657>, (accessed 2018-10-31).
- Piowar, H. A., & Chapman, W. W. (2008). Identifying data sharing in biomedical literature. *AMIA*

- Annual Symposium proceedings, 2008*, 596–600. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655927/>, (accessed 2018-10-31).
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE*, 2(3), e308. Retrieved from <https://doi.org/10.1371/journal.pone.0000308>, (accessed 2018-10-31).
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. Retrieved from <https://doi.org/10.7717/peerj.175>, (accessed 2018-10-31).
- Poldrack, R. A., & Poline, J. (2015). The publication and reproducibility challenges of shared data. *Trends in Cognitive Sciences*, 19(2), 5961. Retrieved from <https://doi.org/10.1016/j.tics.2014.11.008>, (accessed 2018-10-31).
- Price, D. J. (1964). Ethics of scientific publication. *Science*, 144(3619), 655–657. Retrieved from <https://www.jstor.org/stable/1712876>, (accessed 2018-10-31).
- Primary Research Group. (2013). *International survey of academic library data curation practices*. New York, NY: Primary Research Group.
- Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A., & Asgari-Targhi, M. (2010). Adoption and use of Web 2.0 in scholarly communications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4039–4056. Retrieved from <https://doi.org/10.1098/rsta.2010.0155>, (accessed 2018-10-31).
- Prosser, D. C. (2013). Researchers and scholarly communications: an evolving interdependency. In D. Shorley & M. Jubb (Eds.), *The Future of Scholarly Communication* (pp. 39–49). London, United Kingdom: Facet Publishing.
- Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics*, 6(1), 121–130. Retrieved from <https://doi.org/10.1016/j.joi.2011.09.002>, (accessed 2018-10-30).
- Ravez, J. (2018). Opening the heart of science: A review of the changing roles of research libraries. *Publications*, 6(1), 1–13. Retrieved from <https://doi.org/10.3390/publications6010009>, (accessed 2018-10-31).
- Research Data Alliance (RDA) Europe. (2014). The data harvest: How sharing research data can yield knowledge, jobs, and growth. Retrieved from <https://rd-alliance.org/sites/default/files/attachment/The%20Data%20Harvest%20Final.pdf>, (accessed 2018-10-31).
- RECODE project consortium. (2014). *Policy recommendations for Open Access to Research Data*. Retrieved from http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf, (accessed 2015-9-21).
- Renwick, S., Winter, M., & Gill, M. (2017). Managing research data at an academic library in a

- developing country. *IFLA Journal*, 43(1), 51–64. Retrieved from <https://doi.org/10.1177/0340035216688703>, (accessed 2018-10-31).
- Rice, R., & Southall, J. (2016). *The Data Librarian's Handbook*. London, United Kingdom: Facet Publishing.
- Robinson, P. N. (2014). Genomic data sharing for translational research and diagnostics. *Genome Medicine*, 6(9). Retrieved from <https://doi.org/10.1186/s13073-014-0078-2>, (accessed 2018-10-31).
- Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2015). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, 67(12), 29642975. Retrieved from <http://doi.org/10.1002/asi.23529>, (accessed 2018-10-31).
- Roche, D. G., Kruuk, L. E., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLOS Biology*, 13(11), e1002295. Retrieved from <https://doi.org/10.1371/journal.pbio.1002295>, (accessed 2018-10-31).
- Rodwin, M. A., & Abramson, J. D. (2012). Clinical trial data as a public good. *JAMA*, 308(9), 871–872. Retrieved from <https://doi.org/10.1001/jama.2012.9661>, (accessed 2018-10-31).
- Royal Society. (2012). *Science as an open enterprise*. London, United Kingdom: The Royal Society Science Policy Center. Retrieved from <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>, (accessed 2018-10-31).
- Salmi, Jamil. (2014). *Study on Open Science: Impact, implications and policy options*. Luxembourg City, Luxembourg: Publications Office of the European Union. Retrieved from <https://doi.org/10.2777/133494>, (accessed 2018-10-31).
- Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, 80(1), 1–22. Retrieved from <https://doi.org/10.1111/1467-9299.00292>, (accessed 2018-10-31).
- Sanfilippo, P., Hewitt, A. W., & Mackey, D. A. (2018). Plurality in multi-disciplinary research: Multiple institutional affiliations are associated with increased citations. *PeerJ*, 6, e5664. Retrieved from <https://doi.org/10.7717/peerj.5664>, (accessed 2018-10-31).
- satomacoto. (2012). AttachLabelsToPoints.bas [Source code]. Retrieved from <https://gist.github.com/satomacoto/3149548#file-attachlabelstopoints-bas>, (accessed 2018-10-31).
- Sayogo, D. S., Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30, Supplement 1, S19–S31. Retrieved from <https://doi.org/10.1016/j.giq.2012.06.011>, (accessed 2018-10-31).
- Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016a). Open data in global environmental research:

- The Belmont Forum's open data survey. *PLOS ONE*, 11(1), e0146695. Retrieved from <https://doi.org/10.1371/journal.pone.0146695>, (accessed 2018-10-31).
- Schmidt, B., & Shearer, K. (2016b). *Librarians' competencies profile for Research Data Management*. Retrieved from https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf, (accessed 2018-10-31).
- Schofield, P. N., Bubela, T., Weaver, T., Portilla, L., Brown, S. D., Hancock, J. M., . . . Rsenthal, N. (2009). Post-publication sharing of data and tools. *Nature*, 461(7261), 171–173. Retrieved from <https://doi.org/10.1038/461171a>, (accessed 2018-10-31).
- Science. (2011). Challenges and opportunities. *Science*, 331(6018), 692–693. Retrieved from <https://doi.org/10.1126/science.331.6018.692>, (accessed 2018-10-31).
- Scientific Data. (2017). Open for business. *Scientific Data*, 4, 170058(2017). Retrieved from <https://doi.org/10.1038/sdata.2017.58>, (accessed 2018-10-31).
- Scientize. (2014, March 10). Green paper on Citizen Science for Europe: Towards a society of empowered citizens and enhanced research. Retrieved from <https://ec.europa.eu/digital-single-market/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research>, (accessed 2018-10-31).
- Sears, J. R. (2011). *Data sharing effect on article citation rate in paleoceanography*. Paper presented at American Geophysical Union, Fall meeting 2011.
- Sheehan, Jerry. (2017, January 9). *Making federal research results available to all* [Web log post]. Retrieved from <https://obamawhitehouse.archives.gov/blog/2017/01/09/making-federal-research-results-available-all>, (accessed 2018-10-31).
- Shorish, Y. (2012). Data curation is for everyone! The case for master's and baccalaureate institutional engagement with data curation. *Journal of Web Librarianship*, 6(4), 263–273. Retrieved from <https://doi.org/10.1080/19322909.2012.729394>, (accessed 2018-10-31).
- Si, L., Xing, W., Zhuang, X., Hua, X., & Zhou, L. (2015). Investigation and analysis of research data services in university libraries. *The Electronic Library*, 33(3), 417–449. Retrieved from <https://doi.org/10.1108/EL-07-2013-0130>, (accessed 2018-10-31).
- Silvello, G. (2017). Theory and practice of data citation. *Journal of the Association for Information Science and Technology*, 69(1), 6–20. Retrieved from <https://doi.org/10.1002/asi.23917>, (accessed 2018-10-31).
- Smith, A. M., Katz, D. S., Niemeyer, K. E., & FORCE11 Software Citation Working Group. (2016). Software citation principles. *PeerJ Computer Science*, 2, e86. Retrieved from <https://doi.org/10.7717/peerj-cs.86>, (accessed 2018-10-31).
- Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, 41(1), 643–681. Retrieved from <https://doi.org/10.1002/aris.2007.1440410121>, (accessed 2018-10-31).

- Soranno, P. A., Cheruvilil, K. S., Elliott, K. C., & Montgomery, G. M. (2014). It's good to share: Why environmental scientists' ethics are out of date. *BioScience*, 65(1), 69–73. Retrieved from <https://doi.org/10.1093/biosci/biu169>, (accessed 2018-10-31).
- SPARC Europe. (2017a, March 16). Open data and open science policy in Europe. Retrieved from <https://sparceurope.org/open-data-open-science-policy-europe/>, (accessed 2018-10-31).
- SPARC Europe. (2017b, May 31). New SPARC Europe Report provides analysis of open data and open science policies in Europe. Retrieved from <https://sparceurope.org/new-sparc-europe-report-analyses-open-data-open-science-policies-europe/>, (accessed 2018-10-31).
- SPARC Europe, & Digital Curation Centre (DCC). (2017, October). *A snapshot of Open Data and Open Science policies in Europe* (Ver. 2.0). Retrieved from <https://docs.google.com/document/d/1bC7EHsq6yplVKti6HMgKVhaR3T0qfRMwe2oSsej1xs0/edit>, (accessed 2018-10-31).
- SPARC Europe, & Digital Curation Centre (DCC). (2018, January). *An analysis of Open Data and Open Science policies in Europe* (Ver. 2.1). Retrieved from <https://sparceurope.org/download/2285/>, (accessed 2018-10-31).
- Spicer, R. A., & Steinbeck, C. (2018). A lost opportunity for science: Journals promote data sharing in metabolomics but do not enforce it. *Metabolomics*, 14(16). Retrieved from <https://doi.org/10.1007/s11306-017-1309-5>, (accessed 2018-10-31).
- Springer Nature. (2016, December 6). Over 600 Springer Nature journals commit to new data sharing policies. Retrieved from <https://group.springernature.com/in/group/media/press-releases/archive-2016/over-600-springer-nature-journals-commit-to-new-data-sharing-policies/12000254>, (accessed 2018-10-31).
- Springer Nature. (2018). Research data: challenges and opportunities for Japanese researchers. Retrieved from <https://www.springernature.com/gp/open-research/open-data/japan-survey>, (accessed 2018-10-31).
- Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R., Duerr, R., . . . Clark, T. (2015). Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*, 1, e1. Retrieved from <https://doi.org/10.7717/peerj-cs.1>, (accessed 2018-10-31).
- Steele, C. (2013). Recent developments in scholarly communication: A review. *The Australian Library Journal*, 62(3), 224–230. Retrieved from <https://doi.org/10.1080/00049670.2013.831392>, (accessed 2018-10-31).
- Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why has the number of scientific retractions increased? *PLOS ONE*, 8(7), e68397. Retrieved from <https://doi.org/10.1371/journal.pone.0068397>, (accessed 2018-10-31).
- Stodden, V., Guo, P., & Ma, Z. (2013). Toward reproducible computational research: an empirical analysis of data and code policy adoption by journal. *PLOS ONE*, 8(6), e67111. Retrieved

- from <https://doi.org/10.1371/journal.pone.0067111>, (accessed 2018-10-31).
- Strasser, C. (2016). *Research data management 研究データ管理* (機関リポジトリ推進委員会, Trans.). (Original work published 2015). Retrieved from <http://id.nii.ac.jp/1280/00000195/>, (accessed 2018-10-31).
- Sturges, P., Bamkin, M., Anders, J. H., Hubbard, B., Hussain, A., & Heeley, M. (2015). Research data sharing: Developing a stakeholder-driven model for journal policies. *Journal of the Association for Information Science and Technology*, 66(12), 2445–2455. Retrieved from <https://doi.org/10.1002/asi.23336>, (accessed 2018-10-31).
- Suber, P. (2012). *Open access*. Cambridge, MA: MIT Press.
- Swan, A. (2012). *Policy guidelines for the development and promotion of open access*. Paris, France: United Nations Educational, Scientific and Cultural Organization (UNESCO). Retrieved from <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>, (accessed 2018-10-31).
- Taichman, D. B., Sahni, P., Pinborg, A., Peiperl, L., Laine, C., James, A., . . . Backus, J. (2017). Data sharing statements for clinical trials — A requirement of the International Committee of Medical Journal Editors. *New England Journal of Medicine*, 376(23), 2277–2279. Retrieved from <https://doi.org/10.1056/nejme1705439>, (accessed 2018-10-31).
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLOS ONE*, 6(6), e21101. Retrieved from <https://doi.org/10.1371/journal.pone.0021101>, (accessed 2018-10-31).
- Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36(2), 84-90. Retrieved from <https://doi.org/10.1016/j.lisr.2013.11.003>, (accessed 2018-10-31).
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., . . . Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*, 10(8), e0134826. Retrieved from <https://doi.org/10.1371/journal.pone.0134826>, (accessed 2018-10-31).
- Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., . . . Allard, S. (2017). Research data services in European academic research libraries. *LIBER Quarterly*, 27(1), 23-44. Retrieved from <https://doi.org/10.18352/lq.10180>, (accessed 2018-10-31).
- Teplitzky, S. (2017). Open data, [open] access: Linking data sharing and article sharing in the Earth Sciences. *Journal of Librarianship and Scholarly Communication*, 5, eP2150. Retrieved from <https://doi.org/10.1130/abs/2016am-279102>, (accessed 2018-10-31).
- The BMJ. (2012). BMJ Open data campaign. Retrieved from <http://www.bmj.com/open-data>, (accessed 2018-10-31).
- Thomson Reuters. (2013, September 3). Journals list. Retrieved from <http://incites->

- help.isiknowledge.com/incitesLive/ESIGroup/overviewESI/esiJournalsList.html, (accessed 2014-5-8).
- Treloar, A. (2014). The Research Data Alliance: globally co-ordinated action against barriers to data publishing and sharing. *Learned Publishing*, 27(5), S9–S13. Retrieved from <https://doi.org/10.1087/20140503>, (accessed 2018-10-31).
- Valen, D., & Blanchat, K. (2015). *Overview of OSTP responses*. Retrieved from <https://doi.org/10.6084/m9.figshare.1367165.v7>, (accessed 2018-10-31).
- van den Eynden, V., & Bishop, L. (2014). *Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective*. Bristol, United Kingdom: Knowledge Exchange. Retrieved from http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf, (accessed 2018-10-31).
- van Noorden, R. (2014). Confusion over publisher's pioneering open-data rules. *Nature*, 515(7528), 478. Retrieved from <https://doi.org/10.1038/515478a>, (accessed 2018-10-31).
- van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., . . . Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health*, 14(1). Retrieved from <https://doi.org/10.1186/1471-2458-14-1144>, (accessed 2018-10-31).
- Vandewalle, P. (2012). Code sharing is associated with research impact in image processing. *Computing in Science & Engineering*, 14(4), 42–47. Retrieved from <https://doi.org/10.1109/MCSE.2012.63>, (accessed 2018-10-31).
- Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428–436. Retrieved from <https://doi.org/10.1016/j.jbusres.2017.12.043>, (accessed 2018-10-31).
- Vines, T. H., Albert, A. Y., Andrew, R.L., Débarre, F., Bock, D. G., Franklin M. T., . . . Rennison, D. J. (2013a). *The availability of research data declines rapidly with article age* [Data file]. Retrieved from <https://doi.org/10.5061/dryad.q3g37>, (accessed 2018-10-31).
- Vines, T. H., Albert, A. Y., Andrew, R.L., Débarre, F., Bock, D. G., Franklin M. T., . . . Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, 24(1), 94-97. Retrieved from <https://doi.org/10.1016/j.cub.2013.11.014>, (accessed 2018-10-31).
- Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., . . . Yeaman, S. (2013b). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, 27(4), 1304–1308. Retrieved from <https://doi.org/10.1096/fj.12-218164>, (accessed 2018-10-31).
- Vlaeminck, S. (2013). Data management in scholarly journals and possible roles for libraries: Some insights from EDaWaX. *LIBER Quarterly*, 23(1), 48–79. Retrieved from <https://doi.org/10.18352/lq.8082>, (accessed 2018-10-31).

- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., . . . Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14-26. Retrieved from <https://doi.org/10.1016/j.joi.2010.06.004>, (accessed 2018-10-31).
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLOS ONE*, 8(7), e67332. Retrieved from <https://doi.org/10.1371/journal.pone.0067332>, (accessed 2018-10-31).
- Wang, M., & Fong, B. L. (2015). Embedded data librarianship: A case study of providing data management support for a science department. *Science & Technology Libraries*, 34(3), 228–240. Retrieved from <https://doi.org/10.1080/0194262X.2015.1085348>, (accessed 2018-10-31).
- Weller, T., & Monroe-Gulick, A. (2014). Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, 32(3), 467–482. Retrieved from <https://doi.org/10.1108/lht-02-2014-0021>, (accessed 2018-10-31).
- Weingart, P., & Taubert, N. C. (2017). *The future of scholarly publishing Open Access and the economics of digitisation*. Cape Town, South Africa: African Minds.
- Whitlock, M. C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution*, 26(2), 61–65.
- Whyte, A. (2014). *Five steps to decide what data to keep: A checklist for appraising research data* (Ver.1). Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>, (accessed 2018-10-31).
- Whyte, A., & Pryor, G. (2011). Open Science in practice: Researcher perspectives and participation. *International Journal of Digital Curation*, 6(1), 199–213. Retrieved from <https://doi.org/10.2218/ijdc.v6i1.182>, (accessed 2018-10-31).
- Whyte, A., & Sisú, D. (2014). *Results of the Digital Curation Centre 2014 RDM Survey- Briefing 1*. Retrieved from <https://doi.org/10.5281/zenodo.10711>, (accessed 2018-10-31).
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and quality of reporting of statistical results. *PLOS ONE*, 6(11), e26828. Retrieved from <https://doi.org/10.1371/journal.pone.0026828>, (accessed 2018-10-31).
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728.
- Wiley. (2014). Researcher Data Sharing Insights [Infographic]. Retrieved from <https://hub.wiley.com/servlet/JiveServlet/downloadImage/71321/Researcher-Data-Insights-Infographic-FINAL-REVISED-2.jpg>, (accessed 2018-10-31).
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B.

- (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018(2016). Retrieved from <https://doi.org/10.1038/sdata.2016.18>, (accessed 2018-10-31).
- Womack, R. P. (2015). Research data in core journals in Biology, Chemistry, Mathematics, and Physics. *PLOS ONE*, 10(12), e0143460. Retrieved from <https://doi.org/10.1371/journal.pone.0143460>, (accessed 2018-10-31).
- Wuest, T., Mak-Dadanski, J., & Thoben, K. (2014). Data quality in materials science: A quality management manual approach. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications Lecture Notes in Computer Science (Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World)*, 438, 42–49. Retrieved from https://doi.org/10.1007/978-3-662-44739-0_6, (accessed 2018-10-31).
- Zenk-Möltgen, W., & Lepthien, G. (2014). Data sharing in sociology journals. *Online Information Review*, 38(6), 709–722. Retrieved from <https://doi.org/10.1108/OIR-05-2014-0119>, (accessed 2018-10-31).
- Zimmerman, A. S. (2003). *Data sharing and secondary use of scientific data: Experiences of ecologists* (Doctoral Dissertation, University of Michigan). Retrieved from <http://hdl.handle.net/2027.42/61844>, (accessed 2018-10-31).
- Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in sharing and reuse of ecological data. *Science, Technology, & Human Values*, 33(5), 631–652. Retrieved from <https://doi.org/10.1177/0162243907306704>, (accessed 2018-10-31).
- 天野絵理子. (2016). 京都大学オープンアクセス方針の目指すもの. *表面科学*, 37(6), 268–272. Retrieved from <https://doi.org/10.1380/jsssj.37.268>, (accessed 2018-10-31).
- 新たな情報財検討委員会. (2017). 新たな情報財検討委員会報告書：データ・人工知能（AI）の利活用促進による産業競争力強化の基盤となる知財システムの構築に向けて. 知的財産戦略本部検証・評価・企画委員会. Retrieved from http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho.pdf, (accessed 2018-10-31).
- 池内有為. (2013). 研究データ公開に関する学術雑誌のポリシー分析. *三田図書館・情報学会研究大会発表論文集*, 9–12.
- 池内有為. (2014a). 研究データ共有時代における図書館の新たな役割：研究データマネジメントとデータキュレーション. *カレントアウェアネス*, (319), 2126. Retrieved from <https://doi.org/10.11501/8484054>, (accessed 2018-10-31).
- 池内有為. (2014b). 大学図書館による研究データ管理の最前線：研究力を強化するエディンバラ大学の事例. *現代の図書館*, 52(4), 227236. Retrieved from <http://hdl.handle.net/2241/00124096>, (accessed 2018-10-31).

- 池内有為. (2016). オープンサイエンスの論点：Research Data Alliance 第7回総会の議論と日本の趨勢. *情報管理*, 59(4), 241–249. Retrieved from <https://doi.org/10.1241/johokanri.59.241>, (accessed 2018-10-01).
- 池内有為. (2017). オープンサイエンスの展望：公開データの活用事例と課題解決に向けた取り組み. *薬学図書館*, 62(4), 211–217. Retrieved from <http://hdl.handle.net/2241/00150449>, (accessed 2018-10-31).
- 池内有為. (2018a). 日本における研究データ公開の状況と推進要因，阻害要因の分析. *Library and Information Science*, (79), 21–57.
- 池内有為. (2018b). データマネジメントプラン（DMP）：FAIR 原則の実現に向けた新たな展開. *情報の科学と技術*, 68(12), 613–615. Retrieved from https://doi.org/10.18919/jkg.68.12_613, (accessed 2018-12-01).
- 池内有為. (2018c). データ引用：新たな規範への道のり. *情報の科学と技術*, 68(9), 467–469. Retrieved from https://doi.org/10.18919/jkg.68.9_467, (accessed 2018-10-01).
- 池内有為, & 逸村裕. (2016). 学術雑誌によるデータ共有ポリシー：分野間比較と特徴分析. *日本図書館情報学会誌*, 62(1), 20–37. Retrieved from https://doi.org/10.20651/jslis.62.1_20, (accessed 2018-10-31).
- 池内有為, 林和弘, & 赤池伸一. (2017b). 研究データ公開と論文のオープンアクセスに関する実態調査 (NISTEP RESEARCH MATERIAL, 268). 東京, 日本: 科学技術・学術政策研究所. Retrieved from <https://doi.org/10.15108/rm268>, (accessed 2018-10-31).
- 池内健太, 元橋一之, 田村龍一, & 塚田尚稔. (2017a). 科学・技術・産業データの接続と産業の科学集約度の測定 (NISTEP DISCUSSION PAPER, 142). 東京, 日本: 科学技術・学術政策研究所. Retrieved from <https://doi.org/10.15108/dp142>, (accessed 2018-10-31).
- 逸村裕. (2015). オープンアクセス・オープンデータ・オープンサイエンス. 知識と情報. 27(3), 90–95. Retrieved from https://doi.org/10.3156/jsoft.27.3_90, (accessed 2018-10-31).
- 内田治. (2011). *SPSS によるロジスティック回帰分析*. 東京, 日本: オーム社.
- 内田治. (2014). *SPSS によるノンパラメトリック検定*. 東京, 日本: オーム社.
- 大久保嘉子, & 赤井誠. (2005). 引用法とジャーナル・インパクト・ファクターによる研究実績評価の留意点. *科学技術計画*, 20(3), 239–258. Retrieved from https://doi.org/10.20801/jsrpim.20.3_239, (accessed 2018-10-31).
- 大園隼彦. (2017). オープンサイエンスの最新情報：メタデータの相互運用性を中心に. *薬学図書館*, 32(1), 40–47.
- 大向一輝. (2013). オープンデータと Linked Open Data. *情報処理*, 54(12), 1204–1210. Retrieved from <http://id.nii.ac.jp/1001/00095922/>, (accessed 2018-10-31).
- 大波純一, 八塚茂, 信定知江, 箕輪真理, 三橋信孝, & 畠中秀樹. (2018, April 19). データ共有の基準としての FAIR 原則. バイオサイエンスデータベースセンター. Retrieved from <https://doi.org/10.18908/a.2018041901>, (accessed 2018-10-31).

- 岡山将也, & 岩崎一正. (2018). 企業におけるオープンサイエンスの必要性. *情報の科学と技術*, 68(10), 484–488. Retrieved from https://doi.org/10.18919/jkg.68.10_484, (accessed 2018-10-31).
- 小野雅史, 小池俊雄, & 柴崎亮介. (2016). 地球環境情報分野における研究データ共有に関する意識調査：研究現場の実態. *情報管理*, 59(8), 514–525. Retrieved from <https://doi.org/10.1241/johokanri.59.514>, (accessed 2018-10-31).
- 小野寺夏生. (2013). 雑誌インパクトファクターは個人の業績評価に使えない. *現代化学*, (510), 18–22.
- 小野寺夏生. (2016). 論文データベースにおける主題分類：情報分析への利用の視点から. *情報の科学と技術*, 66(6), 272–276. Retrieved from https://doi.org/10.18919/jkg.66.6_272, (accessed 2018-10-31).
- 海洋研究開発機構. (2007, May 16). データ・サンプルの取り扱いに関する基本方針（データポリシー）. Retrieved from https://www.jamstec.go.jp/j/database/data_policy.html, (accessed 2018-10-31).
- 科学技術振興機構. (2017). オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針. 科学技術振興機構. Retrieved from <http://www.jst.go.jp/pr/intro/openscience/>, (accessed 2018-10-31).
- 科学技術振興機構科学技術情報委員会. (2015). わが国におけるデータシェアリングのあり方に関する提言. Retrieved from http://jipsti.jst.go.jp/about/pdf/recommendations_on_data_sharing.pdf, (accessed 2018-10-31).
- 科学技術振興機構戦略研究推進部・研究プロジェクト推進部. (2016). 戦略的創造研究推進事業におけるデータマネジメント実施方針. Retrieved from https://www.jst.go.jp/kisoken/crest/manual/data_houshin.pdf, (accessed 2018-10-31).
- 科学技術・学術政策研究所. (2016). 科学技術の状況に係る総合的意識調査（NISTEP 定点調査 2015）報告書（NISTEP REPORT, 166）. 東京, 日本: 科学技術・学術政策研究所. Retrieved from <https://doi.org/10.15108/nr166>, (accessed 2018-10-31).
- 岸田和明. (2015). 図書館情報学における統計的方法. 東京, 日本: 樹村房.
- 北本朝展. (2015, September 17). オープンサイエンスへのコンバージェンス：同床異夢から共通認識を醸成するコミュニティの形成. オープンサイエンスデータ推進ワークショップ. 京都, 日本.
- 倉田敬子. (2007). 学術情報流通とオープンアクセス. 東京, 日本: 勁草書房.
- 栗原千絵子, & 齊尾武郎. (2017). ICMJE 臨床試験データ共有の声明の背景と今後の課題. *臨床評価*, 45(2), 481–492. http://cont.o.oo7.jp/45_2/p481-92.pdf, (accessed 2018-10-31).
- 経済産業省. (2017a). 「不正競争防止法等の一部を改正する法律案」が閣議決定されました. Retrieved from <http://www.meti.go.jp/press/2017/02/20180227001/20180227001.html>,

- (accessed 2018-10-31).
- 経済産業省. (2017b). *委託研究開発におけるデータマネジメントに関する運用ガイドライン (別冊)*. Retrieved from http://www.meti.go.jp/policy/innovation_policy/data_guideline201712.pdf, (accessed 2018-10-31).
- 国際的動向を踏まえたオープンサイエンスに関する検討会. (2015). *我が国におけるオープンサイエンス推進のあり方について：サイエンスの新たな飛躍の時代の幕開け*. 東京, 日本: 内閣府. Retrieved from <http://www8.cao.go.jp/cstp/sonota/openscience/>, (accessed 2018-10-31).
- 国際動向を踏まえたオープンサイエンスの推進に関する検討会. (2018). *国立研究開発法人におけるデータポリシー策定のためのガイドライン*. 東京, 日本: 内閣府. Retrieved from <https://www8.cao.go.jp/cstp/stsonota/datapolicy/datapolicy.pdf>, (accessed 2018-10-31).
- 国立環境研究所. (2017, April 1). *データの公開に関する基本方針 (データポリシー)*. Retrieved from https://www.nies.go.jp/kihon/kitei/kt_datapolicy.pdf, (accessed 2018-10-31).
- 小谷正雄. (1974). *科学データ：活用と検索*. 東京, 日本: 日本ドクメンテーション協会.
- 近藤康久. (2016). レポート紹介「地球環境研究におけるオープンデータ：ベルモント・フォーラムによるオープンデータ調査」. *情報管理*, 59(4), 250–258. Retrieved from <https://doi.org/10.1241/johokanri.59.250>, (accessed 2018-10-31).
- 近藤康久, & 林和弘. (2018). オープンサイエンスの社会課題解決に対する貢献：マルチステークホルダー・ワークショップによる予測 (NISTEP DISCUSSION PAPER, 163). 東京, 日本: 科学技術・学術政策研究所. Retrieved from <https://doi.org/10.15108/dp163>, (accessed 2018-11-30).
- 佐藤翔. (2016). 査読の抱える問題とその対応策. *情報の科学と技術*, 66(3), 115–121. Retrieved from https://doi.org/10.18919/jkg.66.3_115, (accessed 2018-10-31).
- 佐藤翔. (2018). 学術情報流通と大学図書館：オープンアクセスからオープンサイエンスへ. *図書館界*, 70(1), 245–264.
- 佐藤博樹, 石田浩, & 池田謙一. (2000). *社会調査の公開データ: 2 次分析への招待*. 東京, 日本: 東京大学出版会.
- ジャパンリンクセンター運営委員会. (2015). *研究データへの DOI 登録ガイドライン*. Retrieved from https://doi.org/10.11502/rd_guideline_ja, (accessed 2018-10-31).
- 新エネルギー・産業技術総合開発機構 (NEDO). (2018). *NEDO プロジェクトにおけるデータマネジメントに係る基本方針*. Retrieved from <http://www.nedo.go.jp/content/100875879.pdf>, (accessed 2018-10-31).
- 末吉互. (2012). データベースと著作権. *情報管理*, 55(2), 125–128. <https://doi.org/10.1241/johokanri.55.125>, (accessed 2018-10-31).
- 鈴木淳子. (2016). *質問紙デザインの技法 (第 2 版)*. 京都, 日本: ナカニシヤ出版.

- 総務省統計局. (n.d.). 「データサイエンス・オンライン講座」のご案内. Retrieved from <https://www.stat.go.jp/dss/online.html>, (accessed 2018-10-31).
- 総務省統計局. (2016). 総括編その 4: 研究関係従業者数. In *統計でみる日本の科学技術研究：平成 27 年科学技術研究調査の結果から*. Retrieved from <http://warp.da.ndl.go.jp/info:ndljp/pid/9914563/www.stat.go.jp/data/kagaku/kekka/pamphlet/s-04.htm>, (accessed 2018-10-31).
- 田上博道. (2000). 産学官が一致して取り組む産業技術力強化：産業技術力強化法. *時の法令*, 1624, 6–18.
- デジタルリポジトリ連合 (DRF). (2016). 平成 27 年度機関リポジトリ担当者オンラインワークショップ「研究データから研究プロセスを知る」. Retrieved from <http://drf.lib.hokudai.ac.jp/drif/index.php?onlineworkshop2015>, (accessed 2018-10-31).
- デジタルリポジトリ連合 (DRF). (2017). 平成 28 年 DRF オンラインワークショップ「第 2 回 研究データから研究プロセスを知る」. Retrieved from <http://drf.lib.hokudai.ac.jp/drif/index.php?onlineworkshop2016>, (accessed 2018-10-31).
- 武田英明. (2018). オープンサイエンスの出自とその方向性. *ふみ*, (10), 1–3. Retrieved from https://www.nijl.ac.jp/pages/cijproject/images/fumi_10.pdf, (accessed 2018-10-31).
- 武田英明, 村山泰啓, & 中島律子. (2016). 研究データへの DOI 登録実験. *情報管理*, 58(10), 763–770. Retrieved from <https://doi.org/10.1241/johokanri.58.763>, (accessed 2018-10-31).
- 常川真央, 天野絵里子, 大園隼彦, 西藺由依, 前田翔太, 松本侑子, ... 山地一禎. (2017). 研究データ管理 (RDM) トレーニングツールの構築と展開. *情報知識学会誌*, 27(4), 362–365. Retrieved from https://doi.org/10.2964/jsik_2017_042, (accessed 2018-10-31).
- 内閣府. (2016). *第 5 期科学技術基本計画*. Retrieved from <http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>, (accessed 2018-10-31).
- 内閣府. (2017). *科学技術イノベーション総合戦略 2017*. Retrieved from <https://www8.cao.go.jp/cstp/sogosenryaku/2017/honbun2017.pdf>, (accessed 2018-10-31).
- 内閣府. (2018). *統合イノベーション戦略*. Retrieved from http://www8.cao.go.jp/cstp/tougosenryaku/tougo_honbun.pdf, (accessed 2018-10-31).
- 永崎研宣. (2013). 人文学分野とサイバーインフラストラクチャ：デジタル・ヒューマニティーズにおける現状と課題. *情報の科学と技術*, 63(9), 369–376. Retrieved from https://doi.org/10.18919/jkg.63.9_369, (accessed 2018-10-31).
- 日本医療研究開発機構 (AMED). (2018). データマネジメントプランの提出について. Retrieved from <https://www.amed.go.jp/content/000030140.pdf>, (accessed 2018-10-31).
- 日本学術会議. (1968). *自然科学におけるデータ確立集成活動の推進について (勧告)*. Retrieved from <http://www.scj.go.jp/ja/info/kohyo/05/07-64-k.pdf>, (accessed 2018-10-31).
- 日本学術会議. (2015). *回答：科学研究における健全性の向上について*. Retrieved from <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf>, (accessed 2018-10-31).

- 日本学術会議. (2016). オープンイノベーションに資するオープンサイエンスのあり方に関する提言. Retrieved from <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t230.pdf>, (accessed 2018-10-31).
- 日本学術会議情報学委員会国際サイエンスデータ分科会. (2014). オープンデータに関する権利と義務：本格的なデータジャーナルに向けて（報告）. Retrieved from <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-h140930-3.pdf>, (accessed 2018-10-31).
- 日本学術振興会. (n.d.). オープンアクセス：科学研究費助成事業. Retrieved from https://www.jsps.go.jp/j-grantsinaid/01_seido/08_openaccess/index.html, (accessed 2018-10-31).
- 林和弘. (2015). オープンな情報流通が促進するシチズンサイエンス（市民科学）の可能性：オープンサイエンスをめぐる新しい潮流（その5）. *科学技術動向*, (150), 21–25. Retrieved from <http://hdl.handle.net/11035/3097>, (accessed 2018-10-31).
- 林和弘, & 村山泰啓. (2015). 研究データ出版の動向と論文の根拠データの公開促進に向けて：オープンサイエンスをめぐる新しい潮流（その3）. *科学技術動向*, (148), 4–9. Retrieved from <http://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-STT148J-4.pdf>, (accessed 2018-10-31).
- 福澤尚美. (2016). 日本の研究者はどのようなジャーナルから論文を発表しているのか：オープンアクセスジャーナルに注目して. *STI horizon*, 2(4), 54–59. Retrieved from <https://doi.org/10.15108/stih.00051>, (accessed 2018-10-31).
- 福山樹里. (2015). DataCite：国立図書館×DOI×研究データ. *カレントアウェアネス*, (324), 8–11. Retrieved from <https://doi.org/10.11501/9396324>, (accessed 2018-10-31).
- 船守美穂. (2017). オープンサイエンス推進に関わる学術機関の役割と課題. *情報知識学会誌*, 27(4), 309–322. Retrieved from https://doi.org/10.2964/jsik_2017_034, (accessed 2018-10-31).
- 古川雅子. (2018). MOOC で学ぶオープンサイエンス時代の研究データ管理支援. *図書館雑誌*, 112(10), 665–667.
- 前田知子. (2016). 研究データを対象とした科学技術情報政策：1960年～2015年. *日本図書館情報学会誌*, 62(4), 268–278. Retrieved from https://doi.org/10.20651/jslis.62.4_268, (accessed 2018-10-31).
- 文部科学省. (2014). 研究活動における不正行為への対応等に関するガイドライン. Retrieved from http://www.mext.go.jp/b_menu/houdou/26/08/1351568.htm, (accessed 2018-10-31).
- 文部科学省. (2016). 教育の情報化の推進：プログラミング教育. Retrieved from http://www.mext.go.jp/a_menu/shotou/zyouhou/detail/1375607.htm, (accessed 2018-10-31).
- 文部科学省. (2018). 著作権法の一部を改正する法律案. Retrieved from http://www.mext.go.jp/b_menu/houan/an/detail/1401718.htm, (accessed 2018-10-31).
- 文部科学省アカデミッククラウドに関する検討会. (2012). ビッグデータ時代におけるアカデ

- ミアの挑戦：アカデミッククラウドに関する検討会提言.
http://www.mext.go.jp/b_menu/shingi/chousa/shinkou/027/gaiyou/1323376.htm
- 文部科学省科学技術・学術審議会学術分科会第8期学術情報委員会. (2016). *学術情報のオープン化の推進について（審議まとめ）*. Retrieved from
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm, (accessed 2018-10-31).
- 文部科学省科学技術・学術審議会総合政策特別委員会. (2017). *総合政策特別委員会における第5期科学技術基本計画の実施状況のフォローアップ等に関する審議のとりまとめ*. Retrieved from
http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu22/houkoku/1383213.htm, (accessed 2018-10-31).
- 文部科学省研究開発局環境エネルギー課. (2016). *我が国における地球観測の推進：GEOS10年実施計画*. Retrieved from
http://warp.da.ndl.go.jp/info:ndljp/pid/9779361/www.mext.go.jp/a_menu/kaihatu/kankyousu/ishin/detail/1285005.htm, (accessed 2018-10-31).
- 三菱UFJリサーチ&コンサルティング. (2017). *平成28年度文部科学省委託調査：科学技術イノベーション政策における「政策のための科学」推進事業におけるオープンサイエンスに関する海外動向の調査分析：報告書*. 東京, 日本: 政策研究大学院大学科学技術イノベーション政策研究センター. Retrieved from
https://scirex.grips.ac.jp/resources/archive/170927_872.html, (accessed 2018-10-31).
- 南山泰之. (2015). データジャーナル：研究データ管理の新たな試み. *カレントアウェアネス*, (325), 19–22. Retrieved from <https://doi.org/10.11501/9497651>, (accessed 2018-10-31).
- 三輪哲, & 林雄亮. (2014). *SPSSによる応用多変量解析*. 東京, 日本: オーム社.
- 村山泰啓. (2016). 研究データ共有の国際動向と図書館の役割：RDA第7回総会報告. *カレントアウェアネス*, (328), 10–14. Retrieved from <http://doi.org/10.11501/10020600>, (accessed 2018-10-31).
- 安原通代, & 小野恵理子. (2016). *京都大学若手人材海外派遣事業ジョン万プログラム（職員）：平成27年度図書系職員海外調査研修：米国の大学等高等学術機関におけるオープンサイエンス推進に向けた動向調査*. Retrieved from
<http://hdl.handle.net/2433/209958>, (accessed 2018-10-31).

全研究業績のリスト

1. 学術論文（査読あり）

- (1) 池内有為. (2018). 日本における研究データ公開の状況と推進要因, 阻害要因の分析. *Library and Information Science*, (79), 21–57. Retrieved from <http://lis.mslib.jp/pdf/LIS079021.pdf>, (accessed 2019-09-03).
- (2) 池内有為, & 逸村裕. (2016). 学術雑誌によるデータ共有ポリシー：分野間比較と特徴分析. *日本図書館情報学会誌*, 62(1), 20–37. Retrieved from https://doi.org/10.20651/jslis.62.1_20, (accessed 2018-10-01).
- (3) Lo, P., Dukic, Z., Chiu, Dickson K. W., Ikeuchi, U., Liu, J., & Lu, Y. (2015). Why librarianship? A comparative study between University of Tsukuba, University of Hong Kong, University of British Columbia and Shanghai University. *Australian Academic & Research Libraries*, 46(3), 194–215. Retrieved from <https://doi.org/10.1080/00048623.2015.1059993>, (accessed 2018-10-01).
- (4) 辻慶太, 滝沢伸也, 佐藤翔, 池内有為, 池内淳, 芳鐘冬樹, & 逸村裕. (2013). 図書館の貸出履歴と書誌情報を用いた図書推薦システムの有効性. *図書館界*, 65(4), 253–267. Retrieved from https://doi.org/10.20628/toshokankai.65.4_253, (accessed 2018-10-01).
- (5) 辻慶太, 黒尾恵梨香, 佐藤翔, 池内有為, 池内淳, 芳鐘冬樹, & 逸村裕. (2012). 図書館の貸出履歴を用いた図書推薦システムの有効性検証. *図書館界*, 64(3), 176–189. Retrieved from https://doi.org/10.20628/toshokankai.64.3_176, (accessed 2018-10-01).

2. 学術論文（査読なし）

- (1) 池内有為. (2017). オープンサイエンスの展望：公開データの活用事例と課題解決に向けた取り組み. *薬学図書館*, 62(4), 211–217. Retrieved from <http://hdl.handle.net/2241/00150449>, (accessed 2018-10-01).
- (2) 池内有為. (2016). オープンサイエンスの論点：Research Data Alliance 第7回総会の議論と日本の趨勢. *情報管理*, 59(4), 241–249. Retrieved from <https://doi.org/10.1241/johokanri.59.241>, (accessed 2018-10-01).
- (3) 池内有為. (2015). 研究データ共有の現在：異分野データの統合とデータ引用, 日本のプレゼンス. *情報管理*, 58(9), 673–682. Retrieved from <https://doi.org/10.1241/johokanri58.673>, (accessed 2018-10-01).
- (4) 池内有為. (2014). 大学図書館による研究データ管理の最前線：研究力を強化するエディンバラ大学の事例. *現代の図書館*, 52(4), 227–236. Retrieved from <http://hdl.handle.net/2241/00124096>, (accessed 2018-10-01).
- (5) 池内有為. (2014). 研究データ共有時代における図書館の新たな役割：研究データマネジメントとデータキュレーション. *カレントアウェアネス*, (319), 21–26. Retrieved from <https://doi.org/10.11501/8484054>, (accessed 2018-10-01).

- (6) 逸村裕, & 池内有為. (2013). インパクトファクターの功罪：科学者社会に与えた影響とそこから生まれた歪み. *月刊化学*, 68(12), 32–36. Retrieved from <http://hdl.handle.net/2241/120257>, (accessed 2018-10-01).

3. 学術記事

- (1) 池内有為. (2019). 研究データ管理 (RDM) の目的地と現在地 (連載：オープンサイエンスのいま) . *情報の科学と技術*, 69(3), 125–127. Retrieved from https://doi.org/10.18919/jkg.69.3_125, (accessed 2019-03-01).
- (2) 池内有為. (2018). データマネジメントプラン (DMP) : FAIR 原則の実現に向けた新たな展開 (連載：オープンサイエンスのいま) . *情報の科学と技術*, 68(12), 613–615. Retrieved from https://doi.org/10.18919/jkg.68.12_613, (accessed 2018-12-01).
- (3) 池内有為. (2018). 『ビッグデータ・リトルデータ・ノーデータ』[書評] . *日本図書館情報学会誌*, 64(3), 136–138. Retrieved from https://doi.org/10.20651/jslis.64.3_136, (accessed 2018-10-01).
- (4) 池内有為. (2018). データ引用—新たな規範への道のり (連載：オープンサイエンスのいま) . *情報の科学と技術*, 68(9), 467–469. Retrieved from https://doi.org/10.18919/jkg.68.9_467, (accessed 2018-10-01).
- (5) 池内有為. (2018). 研究データの公開とライセンスの検討状況 (連載：オープンサイエンスのいま) . *情報の科学と技術*, 68(6), 295–297. Retrieved from https://doi.org/10.18919/jkg.68.6_295, (accessed 2018-10-01).
- (6) 池内有為, & 林和弘. (2017). 研究データ公開と論文のオープンアクセスに関する実態調査：オープンサイエンスの課題と展望. *STI Horizon*, 3(4), 27–32. Retrieved from <https://doi.org/10.15108/stih.00106>, (accessed 2018-10-01).
- (7) 池内有為. (2016). 研究データの法的相互運用性：指針と実施のガイドライン. *カレントアウェアネス-E*, (317), E1871. Retrieved from <http://current.ndl.go.jp/e1871>, (accessed 2018-10-01).
- (8) 池内有為. (2016). オープンサイエンスとデータライブラリアン：大学図書館による研究データサービスの事例と示唆. *図書館雑誌*, 110(12), 763–765.
- (9) 池内有為. (2016). 「Research Data Alliance 第7回総会」参加報告. *人文情報学月報*, 57. Retrieved from <http://www.dhii.jp/DHM/dhm57-2>, (accessed 2018-10-01).
- (10) 池内有為. (2016). OECD レポート：オープンサイエンスの実現に向けて. *情報管理*, 58(12), 919–923. Retrieved from <https://doi.org/10.1241/johokanri.58.919>, (accessed 2018-10-01).
- (11) 池内有為. (2016). Making Data Count: データレベルのメトリクス (指標) 開発に向けた調査報告. *情報管理*, 58(11), 844–848. Retrieved from <https://doi.org/10.1241/johokanri.58.844>, (accessed 2018-10-01).
- (12) 池内有為. (2014). 電子書籍サービスで変わる大学図書館の業務と展望<報告>. *カレントアウェアネス-E*, (265), E1600. Retrieved from <http://current.ndl.go.jp/e1600>, (accessed 2018-10-01).

2018-10-01).

- (13) 池内有為. (2014). データが変える研究, 教育, ビジネス, 社会: IDCC14. カレントアウェアネス-E, (257), E1550. Retrieved from <http://current.ndl.go.jp/e1550>, (accessed 2018-10-01).
- (14) 池内有為. (2013). 大学図書館による研究データ公開支援にむけて: 英国調査報告. カレントアウェアネス-E, (245), E1481. Retrieved from <http://current.ndl.go.jp/e1481>, (accessed 2018-10-01).
- (15) 池内有為. (2013). 研究者による研究データ共有の決定要因は何か<文献紹介>. カレントアウェアネス-E, (240), E1452. Retrieved from <http://current.ndl.go.jp/e1452>, (accessed 2018-10-01).
- (16) 池内有為. (2012). 多様化する大学院生のための新たな研究図書館サービス. カレントアウェアネス-E, (234), E1412. Retrieved from <http://current.ndl.go.jp/e1412>, (accessed 2018-10-01).
- (17) 池内有為. (2012). 予算削減? 統計でみる米国の大学図書館 10 年<文献紹介>. カレントアウェアネス-E, (228), E1376. Retrieved from <http://current.ndl.go.jp/e1376>, (accessed 2018-10-01).

4. 書籍

- (1) 池内有為. (2019). 研究データ管理. In 竹内比呂也, 逸村裕, & 佐藤義則編. *新編・変わりゆく大学図書館*. 東京, 日本: 勁草書房. (刊行予定)
- (2) 池内有為. (2013). CMS(16th edition)に則った引用のルール. In 吉村富美子. *英文ライティングと引用の作法: 盗用と言われないための英文指導* (pp. 140–156). 東京, 日本: 研究社.

5. 口頭発表 (査読あり)

- (1) Kondo, Y., Wang, G., Ikeuchi, U., Kano, K., Kumazawa, T., Nakashima, K., Onishi, H., Osawa, T., & Sekino, T. (2018, May). *Open team science: A new team-based research methodology for socio-environmental cases in the open science era*. JpGU 2018. Chiba, Japan.
- (2) Ikeuchi, U., Harada, T., Sato, S., Okabe, Y., & Itsumura, H. (2017, September). *Data literacy perceptions and Research Data Management practices by researchers in Japan*. European Conference on Information Literacy (ECIL) 2017. Saint-Malo, France. Retrieved from <http://hdl.handle.net/2241/00148306>, (accessed 2018-10-01).
- (3) Lo, P., Duckic, D., Ikeuchi, U., Liu, J., & Lu, Y. (2014, April). *Why librarianship: A comparative study between University of Tsukuba, University of Hong Kong, University of British Columbia, and Shanghai University*. The Asian Conference on Literature & Librarianship (LibrAsia) 2014. Osaka, Japan.

- (4) Tsuji, K., Takizawa, N., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Itsumura, H. (2013, September). Book Recommendation based on Library Loan Records and Bibliographic Information. *Procedia: Social and Behavioral Sciences*, 147, 478–486. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1877042814040531>, (accessed 2018-10-01).
- (5) Tsuji, K., Kuroo, E., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Itsumura, H. (2012, September). Use of library loan records for book recommendation. *Proceedings of the 3rd IIAI International Conference on e-Services and Knowledge Management (IIAI ESKM 2012)*, 2012, 30–35. Retrieved from <https://doi.org/10.1109/IIAI-AAI.2012.16>, (accessed 2018-10-01).
- (6) Ikeuchi, A., Tsuji, K., Yoshikane, F., & Ikeuchi, U. (2012, August). Double-bounded dichotomous choice CVM for public library services in Japan. *Procedia: Social and Behavioral Sciences*, 73, 205–208. Retrieved from <https://doi.org/10.1016/j.sbspro.2013.02.042>, (accessed 2018-10-01).
- (7) Tsuji, K., Kuroo, E., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Itsumura, H. (2011, October). Use of library loan records for book recommendation. *Proceedings of the International Conference on Integrated Information (IC-ININFO 2011)*, 2011, 4. Retrieved from <https://doi.org/10.1109/IIAI-AAI.2012.16>, (accessed 2018-10-01).

6. 口頭発表（査読なし）

- (1) 池内有為. (2013, November). 研究データ公開に関する学術雑誌のポリシー分析. 三田図書館・情報学会 2013 年度研究大会. 東京, 日本.
- (2) 西野祐子, 松野渉, 池内有為, 佐藤翔, & 逸村裕. (2013, October). 大学図書館における未貸出図書の分析. 第 61 回日本図書館情報学会研究大会. 東京, 日本.
- (3) 池内淳, 辻慶太, 芳鐘冬樹, & 池内有為. (2012, November). 表明選好法による公立図書館の経済価値の測定に関する研究 : CVM とコンジョイント分析の比較. 第 60 回日本図書館情報学会 1 研究大会. 福岡, 日本.
- (4) 松野渉, 本田咲美, 池内有為, 佐藤翔, & 逸村裕. (2012, May). 大学附属図書館における貸出履歴の分析. 2012 年日本図書館情報学会春季研究集会. 三重, 日本.
- (5) 池内淳, 安形輝, 前田美紀, 藤井有為, & 上田修一. (1995, November). 図書館学, 情報学における基礎文献とその通時的引用パターンの調査. 三田図書館・情報学会 1995 年度研究大会. 東京, 日本.

7. ポスター発表（査読あり）

- (1) Kondo, Y., Wang, G., Ikeuchi, U., Kano, K., Kumazawa, T., Nakashima, K., Onishi, H., Osawa, T., & Sekino, T. (2018). Information asymmetry reduction in open team science: Call for international collaborators. BOUNDARY SPANNING: Advances in Socio-Environmental Systems Research, The National Socio-Environmental Synthesis Center (SESYNC),

Maryland, June 12, 2018.

- (2) Kondo, Y., Hayashi, K., Ikeuchi, U., Kuribayashi, M., Yano, S., & Kitamoto, A. (2017). Future of Open Science foreseen with society: report on a multi-stakeholder workshop in Japan. JpGU-AGU Joint Meeting 2017. Chiba, May 23, 2017.
- (3) Ikeuchi, U., & Itsumura, H. (2015). A snapshot of data citation using the Data Citation Index and the Web of Science. 10th International Digital Curation Conference (IDCC 2015). London, February 9-12, 2015.
- (4) Ikeuchi, U., Nishiura, M., & Itsumura, H. (2014). A review of journal policies for sharing research data across disciplines. 9th International Digital Curation Conference (IDCC 2014). San Francisco, February 24-27, 2014.

8. ポスター発表（査読なし）

- (1) 和田匡路, 遠藤裕子, & 池内有為. (2015). 研究データ共有に関わるステークホルダーの動向と課題. 情報メディア学会研究大会. 京都, 2015 年 6 月 27 日.
- (2) 池内有為., Dukic, D., Liu, J., Lu, Y., Xu, Y., & Lo, P. (2013). Why Librarianship? 図書館情報学大学院生の進学要因とキャリア形成の国際比較. 情報メディア学会研究大会. 神奈川, 2013 年 6 月 29 日.

9. 講演, パネルディスカッション

- (1) 池内有為. (2018). オープンサイエンス：情報と公共図書館の接点. 館長セミナー 2018：公共図書館基礎講座. 大阪, 図書館流通センター（TRC）. 2018 年 9 月 28 日.（講演）
- (2) 池内有為. (2018). オープンサイエンス：情報と公共図書館の接点. 館長セミナー 2018：公共図書館基礎講座. 東京, 図書館流通センター（TRC）. 2018 年 9 月 13 日.（講演）
- (3) 池内有為. (2018). 大学図書館と研究支援. 平成 30 年度大学図書館職員長期研修. 茨城, 筑波大学. 2018 年 7 月 12 日.（講演）
- (4) 池内有為. (2018). 研究データ公開におけるライセンスの現状と課題：インタビュー・アンケート調査の結果から. Japan Open Science Summit 2018 (JOSS 2018)：政策・ポリシー「研究データのライセンス条件を考える：産官学ラウンドテーブル」. 東京, 学術総合センター. 2018 年 6 月 18 日.（講演, パネルディスカッション）
- (5) 池内有為. (2018). NISTEP 調査からみた研究者のニーズと支援の可能性. Japan Open Science Summit 2018 (JOSS 2018)：図書館関連「研究データ管理のためのトレーニングコース」. 東京, 学術総合センター. 2018 年 6 月 18 日.（講演）
- (6) 池内有為. (2017). オープンサイエンスの動向と大学図書館への期待. 平成 29 年度東北地区大学図書館協議会合同研修会. 2017 年 8 月 25 日.（講演）
- (7) 池内有為. (2017). 大学図書館と研究支援. 平成 29 年度大学図書館職員長期研修. 茨城, 筑波大学. 2017 年 7 月 13 日.（講演）

- (8) 池内有為. (2017). オープンサイエンスの展望：研究データのオープン化と利活用，イノベーションの創出. 日本薬学会第 137 年会：一般シンポジウム S35 「ヘルスケアの最前線：新しい情報技術を現場に生かす」. 宮城, 仙台国際センター. 2017 年 3 月 26 日. (講演, パネルディスカッション)
- (9) 池内有為. (2017). オープンサイエンスの実現に向けた研究データ共有の実態調査. NISTEP-地球研-NII 合同ワークショップ「社会との協働が切り拓くオープンサイエンスの未来」. 京都, 総合地球環境学研究所. 2017 年 1 月 27-28 日. (講演, ファシリテーター)
- (10) 池内有為. (2016). オープンサイエンスの種をまく：大学図書館とデータライブラリアンの挑戦. 北海道大学オープンサイエンス元年：第 2 回 国内外のオープンサイエンスに関する動向と展望. 北海道, 北海道大学. 2016 年 11 月 4 日. (講演, パネルディスカッション)
- (11) 池内有為. (2016). 進化する研究図書館：研究支援サービスの事例とエッセンス. 平成 28 年度総合研究大学院大学附属図書館研修会. 東京, 大学共同利用機関法人情報・システム研究機構国立極地研究所. 2016 年 10 月 14 日. (講演)
- (12) 池内有為. (2016). オープンサイエンスのキーワードと図書館の戦略, 医学分野の動向. 第 420 回 Reference Conference. 東京, 東邦大学医学メディアセンター. 2016 年 8 月 19 日. (講演)
- (13) 池内有為. (2016). データライブラリアンとその育成：ラーニングコモンズと Digital Scholarship, オープンサイエンス. 日本図書館協会図書館情報学教育部会 2016 年度第 1 回研究集会. 東京, 日本図書館協会. 2016 年 6 月 5 日. (基調講演)
- (14) 池内有為. (2016). RDA の活動と議論, 日本の課題解決への示唆：人材育成と法的枠組み. 研究データとオープンサイエンスフォーラム～RDA 東京大会における議論を踏まえた研究データ共有の最新動向～. 東京, 国立国会図書館. 2016 年 3 月 17 日. (講演)
- (15) 逸村裕, & 池内有為. (2015). オープンサイエンスによる学術情報流通の変化と大学図書館の研究データ管理：Data Citation Index の可能性. 第 17 回図書館総合展「オープンサイエンス時代の研究情報データベース」神奈川, パシフィコ横浜. 2015 年 11 月 11 日. (講演)
- (16) 池内有為. (2015). データライブラリアンに求められる能力と人材育成. NII 国立情報学研究所 学術情報基盤オープンフォーラム 2015. 東京, 学術総合センター. 2015 年 6 月 11 日. (講演)
- (17) 池内有為. (2015). 研究データ共有の意義と展望. 三田図書館・情報学会月例会. 東京, 慶應義塾大学. 2015 年 3 月 7 日. (講演)
- (18) 池内有為. (2014). インパクトファクターについて知っておくべき 10 のことから. 平成 26 年度茨城県図書館協会大学図書館部会研修会「図書館員のための『研究評価』指標の基礎知識：インパクトファクターからオルトメトリクスへ」. 茨城, 筑波大学. 2014 年 11 月 27 日. (講演)

- (19) 池内有為. (2014). オープンサイエンスと大学図書館：研究データ共有の動向と課題. 第 46 回デジタル図書館ワークショップ / 情報処理学会 第 116 回 IFAT 研究発表会 合同研究会. 東京, 筑波大学東京キャンパス文京校舎. 2014 年 11 月 11 日. (特別講演)
- (20) 池内有為. (2014). オープンサイエンスを支えるデータライブラリアン. 第 16 回図書館総合展「大学の知の発信システムの構築に向けて - 機関リポジトリの新たな可能性を探る, 2. 研究データへのアプローチ：エジンバラ大学図書館の実践事例を中心に」. 神奈川, パシフィコ横浜. 2014 年 11 月 6 日. (講演, パネルディスカッション)
- (21) Ikeuchi, U. (2014). Challenges of 3 University Libraries in Japan. Elsevier APAC eBooks Forum 2014, "Challenging the 'norm': Future Directions of eBooks". Brisbane, Australia. June 26-27, 2014. (講演)
- (22) 池内有為. (2014). 研究データマネジメント：データサイエンス時代の大学図書館. 国立大学図書館協会 関東甲信越地区協会セミナー「Beyond Library, Beyond Librarian」. 茨城, 筑波大学. 2014 年 2 月 20 日. (講演)
- (23) 池内有為. (2013). 英国における研究データ管理支援の動向. 第 3 回 SPARC Japan セミナー2013「オープンアクセス時代の研究成果のインパクトを再定義する：再利用と Altmetrics の現在」. 東京, 国立情報学研究所. 2013 年 10 月 25 日. (講演, パネルディスカッション)

10. 受賞

- (1) 10th International Digital Curation Conference (IDCC15) Best Poster Award. 2015 年 2 月 10 日.
- (2) 三田図書館・情報学会 2013 年度研究大会ベストプレゼンテーション賞. 2013 年 11 月 9 日.

11. その他

- (1) 池内有為, 林和弘, & 赤池伸一. (2017). *研究データ公開と論文のオープンアクセスに関する実態調査* (NISTEP RESEARCH MATERIAL No. 268). 東京, 日本: 文部科学省 科学技術・学術政策研究所. Retrieved from <https://doi.org/10.15108/rm268>, (accessed 2018-10-01).
- (2) 池内有為. (2015, March 23). 大学図書館が担う, 分野を超えた知の可能性 [筑波大学 オープンコースウェア (OCW)] . Retrieved from <http://ocw.tsukuba.ac.jp/discovery/researchers/ikeuchi/>, (accessed 2018-10-01).

付録

[付録 1] 先行研究によるデータ公開の動機と障壁の出典

付表 1 先行研究によるデータ公開の動機と障壁の出典

要因	出典
学術雑誌のポリシー	Huang et al., 2012; Wallis, 2013; van den Eynden & Bishop, 2014; Kim & Burns, 2015; Kim & Stanton, 2015; Schmidt et al., 2016a
助成機関のポリシー	Wallis, 2013; Fecher et al., 2015; Kim & Burns, 2015; Schmidt et al., 2016a
所属機関のポリシー	Fecher et al., 2015; Kim & Stanton, 2015; Schmidt et al., 2016a
規範	Ferguson 2014; van den Eynden & Bishop, 2014; Fecher et al. 2015; Kim & Stanton, 2015; Schmidt et al., 2016a; Berghmans et al., 2017
業績	Tenopir et al., 2011; van den Eynden & Bishop, 2014; Ferguson 2014; Tenopir et al., 2015; Kim & Stanton, 2015; Fecher et al., 2015
科学の発展	Damvad, 2014; van den Eynden & Bishop, 2014; Schmidt et al., 2016a; 小野ら, 2016
他の研究者からの要求	Wallis, 2013; Ferguson, 2014; van den Eynden & Bishop, 2014; Kim & Stanton, 2015; Schmidt et al., 2016a
研究成果の認知度向上	van den Eynden & Bishop, 2014; Schmidt et al., 2016a
オープンデータへの貢献	van den Eynden & Bishop, 2014; Schmidt et al., 2016a
分野リポジトリ	Tenopir et al., 2011; van Noorden, 2014; Kim & Zhang, 2015
機関リポジトリ	Tenopir et al., 2011; Wallis, 2013; van Noorden, 2014; Kim & Zhang, 2015
資金	Tenopir et al., 2011; Science, 2011
時間	Tenopir et al., 2011; Damvad, 2014; Kim & Stanton, 2015; 小野ら, 2016
人材	van den Eynden & Bishop, 2014; Kim & Stanton, 2015; 小野ら, 2016; Berghmans et al., 2017
先に論文を出版される可能性	Huang et al., 2012; Damvad, 2014; van Noorden, 2014; van den Eynden & Bishop, 2014; Fecher et al., 2015; Tenopir et al., 2015; 小野ら, 2016; Schmidt et al., 2016a;
誤解や誤用の可能性	Damvad, 2014; Fecher et al., 2015; Tenopir et al., 2015; Schmidt et al., 2016a
引用せずに利用される可能性	Fecher et al., 2015; Tenopir et al., 2015; Schmidt et al., 2016a
研究の誤りを発見される可能性	Wicherts, 2011; van den Eynden & Bishop, 2014; Schmidt et al., 2016a
機密・プライバシー情報	Ferguson, 2014; Fecher et al., 2015
商用利用される可能性	Fecher et al., 2015; Schmidt et al., 2016a
知的財産権	Tenopir et al., 2011; Ferguson, 2014; Fecher et al., 2015; Schmidt et al., 2016a

[付録 2] 調査票

1. 研究分野について

Q1. ご自身の研究分野に最も近いものをお選び下さい。＊（回答必須）

- | | | |
|--|-----------------|----------|
| (1)工学 | (2)天文学 | (3)化学 |
| (4)物理学 | (5)地球科学 | (6)数学 |
| (7)コンピュータサイエンス | (8)農学 | (9)生物科学 |
| (10)医学 | (11)心理学 | (12)社会科学 |
| (13)人文学 | (14)その他〔具体的に： 〕 | |
| (15)研究を行ったことはない（口頭発表や論文出版の経験はない）【→Q35 へ】 | | |

2. 学術論文について

ご自身の論文のオープンアクセス（公開）状況や、論文の利用についてお伺いします。

なお、本調査における「オープンアクセス」とは、論文がインターネットで公開され、読者は無料で読むことができる状態とします。

【オープンアクセスの例】

- ・ オープンアクセスの雑誌で出版する（PLOS など）
- ・ 雑誌のオープンアクセスオプションを選択する（Springer Open Choice など）
- ・ 雑誌等が一定期間経過後に論文をオープンアクセスにする
- ・ 機関リポジトリやプレプリントサーバで論文を公開する（arXiv など）

Q2. ご自身の論文について、あてはまるものをお選び下さい。＊

- (1) オープンアクセスの論文がある
- (2) オープンアクセスの論文はない【→Q4 へ】
- (3) わからない【→Q4 へ】

Q3. 論文をオープンアクセスにした理由として、あてはまるものをお選び下さい。〔複数選択可〕＊

- (1) 他の研究者からのリクエストに応じて
- (2) 論文を投稿した雑誌がオープンアクセスだから
- (3) 助成機関のポリシー（助成条件）だから
- (4) 所属機関のポリシーだから
- (5) 分野・コミュニティの規範だから
- (6) 研究成果を広く認知してもらいたいから
- (7) 科学研究や成果実装を推進したいから

- (8) オープンアクセスに貢献したいから
- (9) その他〔具体的に： 〕
- (10) あてはまるものはない【※黄色マーカーの回答は他の選択肢とは同時に選べません】
【すべて→Q6 へ】

Q4. 論文がオープンアクセスではない理由として、あてはまるものをお選び下さい。〔複数選択可〕＊

- (1) 資金が必要だから
- (2) 時間が必要だから
- (3) ニーズがないと思うから
- (4) 投稿したい雑誌がオープンアクセスではないから
- (5) 助成機関のポリシー（助成条件）ではないから
- (6) 所属機関にオープンアクセス方針がないから
- (7) 分野・コミュニティで推奨されていないから
- (8) リポジトリなどの公開手段がないから
- (9) その他〔具体的に： 〕
- (10) あてはまるものはない【→Q6 へ】

Q5. Q4 の理由が解決された場合、論文をオープンアクセスにしたいと思われますか？
＊

- (1) はい
- (2) いいえ
- (3) わからない

Q6. 研究に利用したことがある（アイデアの参考にしたり引用した経験がある）論文の分野をお選び下さい。〔複数選択可〕＊

- | | | |
|----------------|-----------------|----------|
| (1)工学 | (2)天文学 | (3)化学 |
| (4)物理学 | (5)地球科学 | (6)数学 |
| (7)コンピュータサイエンス | (8)農学 | (9)生物科学 |
| (10)医学 | (11)心理学 | (12)社会科学 |
| (13)人文学 | (14)その他〔具体的に： 〕 | |

Q7. 論文を探す際に、よく利用する検索ツールや情報源をお選び下さい。〔複数選択可〕

- (1) サーチエンジン（Google, Google Scholar など）
- (2) 論文情報のデータベース（Web of Science, Scopus, CiNii Articles など）
- (3) 出版社や学術雑誌のサイト（Elsevier, Wiley など）
- (4) 学術機関のリポジトリ・アーカイブ（大学やNASAの機関リポジトリなど）
- (5) 論文やプレプリントのサーバ（PubMed Central, arXiv, J-Stage など）

- (6) 論文や学術記事の参考文献
- (7) 政府・機関・出版社などの広報、ニューズレター
- (8) ブログや一般的な SNS (Facebook, Twitter など)
- (9) 学術系 SNS (Mendeley, ResearchGate など)
- (10) アラートサービス (RSS など)
- (11) メーリングリスト
- (12) 研究者や同僚に尋ねる／教えてもらう
- (13) その他 [具体的に：]

Q8. 論文を利用する際に、その信頼性の判断基準としている項目をお選び下さい。[複数選択可]

- (1) 著者情報 (所属機関、職位など)
- (2) 掲載雑誌
- (3) 引用数 (その論文を引用した文献数)
- (4) オルトメトリクス (SNS やブログの言及数)
- (5) ダウンロード数
- (6) 抄録
- (7) 研究手法の確かさ
- (8) その他 [具体的に：]

3. 研究データの提供について

本調査における「データ」とは、研究のために収集・作成・観測したデジタルデータを指します。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しません。また、ゲノムデータ、地理情報、ソフトウェアコード、インタビューの録音と書き起こしなど、内容も限定しません。

データの「提供」とは、E-mail や USB フラッシュメモリ、クラウドサービス (Dropbox や Google Drive) などを使って、共同研究者を除く他者に渡す (共有する) ことを指します。特定の人以外はアクセスできない状態です。

Q9. 共同研究者を除く他の研究者にデータを提供したご経験はありますか？

- (1) よくある
- (2) たまにある
- (3) ほとんどない
- (4) まったくない
- (5) わからない

Q10. 共同研究者を除く他の研究者からデータの提供を受けたご経験はありますか？

- (1) よくある
- (2) たまにある
- (3) ほとんどない
- (4) まったくない
- (5) わからない

4. 公開データの利用について

「公開データ」とは、ウェブサイトやリポジトリ、論文の補足資料などに掲載され、インターネットでアクセスして利用できるデータを指します。利用料金や利用者登録が必要な場合も含まれます。

Q11. これまでに、公開データを以下の公開先から入手したご経験はありますか？ [複数選択可] *

- (1) 個人や研究室のウェブサイト
- (2) 学術機関のリポジトリ・データアーカイブ（大学や NASA のリポジトリなど）
- (3) 特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）
- (4) データ共有サービス（figshare, zenodo など）
- (5) コード共有サービス（GitHub など）
- (6) 論文の補足資料（supplementary materials）
- (7) 学術系 SNS（Mendeley, ResearchGate など）
- (8) その他 [具体的に：]
- (9) **ない【→Q17 へ】**
- (10) **わからない【→Q17 へ】**

Q12. 公開データの入手にあたって、問題だと感じたことがある項目をお選び下さい [複数選択可] *

- (1) 利用料金が必要
- (2) 利用者登録が必要
- (3) リクエストから入手までに時間がかかる
- (4) 最新のデータを入手できない
- (5) データごとにフォーマットが異なる
- (6) データごとに品質が異なる
- (7) データの解釈や再利用方法がよくわからない
- (8) 利用したいデータへのアクセス方法がよくわからない
- (9) 利用条件がよくわからない

- (10) 著作者情報がよくわからない
 (11) その他 [具体的に：]
 (12) 問題を感じたことはない

Q13. 入手した公開データの利用目的について、あてはまるものをそれぞれお選び下さい。

＊

	行わない	たまに行う	行う
自身の研究のアイデアや仮説の参考にする	1	2	3
再分析・再利用して自身の研究を行う	1	2	3
研究を再現・追試する	1	2	3

Q14. 研究に利用したことがある（参考にしたり再分析した経験がある）公開データの分野をお選び下さい。[複数選択可] ＊

- (1)工学 (2)天文学 (3)化学
 (4)物理学 (5)地球科学 (6)数学
 (7)コンピュータサイエンス (8)農学 (9)生物科学
 (10)医学 (11)心理学 (12)社会科学
 (13)人文学 (14)その他 [具体的に：]
 (15)公開データを入手しても、研究に利用したことはない【→Q17へ】

Q15. 公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい。[複数選択可]

- (1) サーチエンジン（Google など）
 (2) データ情報のデータベース（Data Citation Index など）
 (3) 学術機関のリポジトリ・データアーカイブ（大学やNASAのリポジトリなど）
 (4) 出版社や学術雑誌のサイト（Elsevier, Wiley など）
 (5) 特定のデータリポジトリ [具体的に：]
 (6) 論文や学術記事の参考文献
 (7) データジャーナル（簡易なデータ記述とデータへのリンクを掲載した雑誌）
 (8) 政府・機関・出版社などの広報，ニュースレター
 (9) ブログや一般的な SNS（Facebook, Twitter など）
 (10) 学術系 SNS（Mendeley, ResearchGate など）
 (11) アラートサービス（RSS など）
 (12) メーリングリスト

(13) 研究者や同僚に尋ねる／教えてもらう

(14) その他〔具体的に： 〕

Q16. 公開データを利用する際に、信頼性の判断基準としている項目をお選び下さい。〔複数選択可〕

(1) 著者情報（所属機関、職位など）

(2) 掲載リポジトリ

(3) 引用数（そのデータを引用した文献数）

(4) オルトメトリクス（SNS やブログの言及数）

(5) ダウンロード数

(6) 研究手法の確かさ

(7) そのデータを用いた論文

(8) データにつけられた説明（Readme ファイルなど）

(9) その他〔具体的に： 〕

Q17. 今後、利用してみたいと思う公開データの分野をお選び下さい。〔複数選択可〕 ＊

(1)工学

(2)天文学

(3)化学

(4)物理学

(5)地球科学

(6)数学

(7)コンピュータサイエンス

(8)農学

(9)生物科学

(10)医学

(11)心理学

(12)社会科学

(13)人文学

(14)その他〔具体的に： 〕

(15)他の研究者の公開データを利用したいとは思わない

5. 研究データの公開について

これまでに、ご自身の研究データをインターネットで公開したご経験についてお伺いします。

（再掲）本調査における「データ」とは、研究のために収集・作成・観測したデジタルデータを指します。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しません。また、ゲノムデータ、地理情報、ソフトウェアコード、インタビューの録音と書き起こしなど、内容も限定しません。

Q18. これまでに、研究データを以下の方法で公開したご経験はありますか？〔複数選択可〕 ＊

(1) 個人や研究室のウェブサイト

(2) 所属機関のリポジトリ・データアーカイブ

(3) 特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）

- (4) データ共有サービス (figshare, zenodo など)
- (5) コード共有サービス (GitHub など)
- (6) 論文の補足資料 (supplementary materials)
- (7) 学術系 SNS (Mendeley, ResearchGate など)
- (8) その他 [具体的に:]
- (9) 公開したことはない 【→Q20 へ】
- (10) わからない 【→Q20 へ】
- (11) 研究にデジタルデータは用いない 【→Q35 へ】

Q19. 研究データを公開した理由として、あてはまるものをお選び下さい。[複数選択可]

＊

- (1) 他の研究者からのリクエストに応じて
- (2) 論文を投稿した雑誌のポリシー (投稿規定) だから
- (3) 助成機関のポリシー (助成条件) だから
- (4) 所属機関のポリシーだから
- (5) 分野・コミュニティの規範だから
- (6) 業績になる場合があるから
- (7) 研究成果を広く認知してもらいたいから
- (8) 科学研究や成果実装を推進したいから
- (9) オープンデータに貢献したいから
- (10) その他 [具体的に:]
- (11) あてはまるものはない
【すべて→Q22 へ】

Q20. 研究データを公開していない理由として、あてはまるものをお選び下さい。[複数選択可] ＊

- (1) 資金が必要だから
- (2) 時間が必要だから
- (3) ニーズがないと思うから
- (4) 論文を投稿した雑誌のポリシー (投稿条件) ではないから
- (5) 助成機関のポリシー (助成条件) ではないから
- (6) 所属機関にデータ公開ポリシーがないから
- (7) 分野・コミュニティで推奨されていないから
- (8) 業績にならないから
- (9) リポジトリなどの公開手段がないから
- (10) その他 [具体的に:]
- (11) あてはまるものはない 【→Q22 へ】

Q21.Q20 の理由が解決された場合、研究データを公開したいと思われますか？＊

- (1) はい
- (2) いいえ
- (3) わからない

Q22.所属機関では、研究データの保存期間は定められていますか？

- (1) 定められている
- (2) 定められていない
- (3) わからない

6. 最近の研究のためのデータについて

ここからは、研究データの整備や公開に必要な支援について検討するための質問です。

研究ごとに扱うデータの種類や量が異なる場合が多いと考えられるため、”論文などの成果を発表済みの、最近の主要な研究 1 件のために収集・作成・観測したデータ”についてお伺いします（以下、「カレントデータ」と記します）。

データ公開のご経験がない場合は、カレントデータのうち、論文等の根拠となる部分を公開すると想定してお答え下さい。

【例：研究 1 件のために質問紙調査とインタビュー調査を実施した場合】

＜カレントデータ＞

- ・質問紙の回答を入力したスプレッドシート
- ・回答を分析するための R のコード
- ・インタビューを録音した音声データ
- ・インタビューを書き起こしたテキストデータなど

＜データ公開＞

質問紙調査の回答データを第三者が再利用できるように、調査概要の説明を作成、回答を入力したスプレッドシートから個人情報を削除、項目に見出しをつけるといった処理を行った上で、質問紙や分析のためのコードとともにリポジトリに登録する。

Q23.カレントデータの総量は、およそどれくらいでしたか？ 論文などには使用しなかったデータも含めてあてはまる単位をお選びください。

- (1) MB（メガバイト）以下
- (2) GB（ギガバイト）

- (3) TB（テラバイト）
- (4) PB（ペタバイト）以上
- (5) わからない

Q24. ご自身以外で、カレントデータの所有権をもつ人・組織をお選び下さい。[複数選択可]

- (1) 共同研究者
- (2) 所属機関
- (3) 所属機関以外の研究機関
- (4) 所属機関以外の企業
- (5) 研究助成機関
- (6) その他〔具体的に： 〕
- (7) なし（自分のみ）
- (8) わからない

Q25. カレントデータには、以下の情報が含まれていますか？ [複数選択可]

- (1) 個人情報
- (2) 健康情報（遺伝情報、医療情報）
- (3) 企業・商業上の機密情報
- (4) その他、守秘義務がある情報
- (5) なし
- (6) わからない

Q26. カレントデータは何年くらい保存する必要があると思われますか？ 年数を数字で入力して下さい。（保存の必要はないと考える場合は「0」、永久保存の場合は「999」と入力して下さい）

[] 年

Q27. 現在までに、カレントデータ（データの一部）を公開しましたか？＊

- (1) はい：以下の質問は、（カッコ）をお読みの上お答え下さい
- (2) いいえ：以下の質問は、公開することを想定してお答え下さい
- (3) わからない：以下の質問は、公開することを想定してお答え下さい

Q28. カレントデータを管理・公開しようとする場合、次の資源は十分に整っていますか（いましたか）？＊

	わからない	不十分	やや不十分	ほぼ充分	充分
研究中のデータ用ストレージ	0	1	2	3	4
データ公開用のリポジトリ	0	1	2	3	4
研究終了後のデータ保存用ストレージ	0	1	2	3	4
データの整備・公開のための時間	0	1	2	3	4
データの整備・公開のための人材	0	1	2	3	4
データの整備・公開のための資金	0	1	2	3	4

Q29. カレントデータを論文の発表前に公開しようとする場合、次の点は問題（懸案）となりますか（なりましたか）？＊

	わからない	問題ではない	あまり問題ではない	やや問題である	問題である
商用利用される可能性	0	1	2	3	4
機密・プライバシー情報	0	1	2	3	4
誤解や誤用の可能性	0	1	2	3	4
引用せずに利用される可能性	0	1	2	3	4
研究の誤りを発見される可能性	0	1	2	3	4
公開したデータを使って自分より先に論文を出版される可能性	0	1	2	3	4

Q30. カレントデータを公開した場合（公開したカレントデータは）、ご自身と同じ分野、すなわち Q1 で選択した分野の研究者の多くが理解できると思われますか？＊

- (1) できると思う：データにつけた説明（Readme ファイルなど）を読めば理解できると思う
- (2) やや難しいと思う
- (3) 難しいと思う
- (4) わからない

Q31. カレントデータを公開した場合（公開したカレントデータは）、異分野の研究者の多くが理解できると思われますか？＊

- (1) できると思う：データにつけた説明（Readme ファイルなど）を読めば理解できると思う
- (2) やや難しいと思う
- (3) 難しいと思う
- (4) わからない

Q32. カレントデータを公開した場合（公開したカレントデータについて）、ご関心がある項目をお選び下さい。[複数選択可]

- (1) 閲覧数
- (2) ダウンロード数
- (3) リンク数（別のウェブサイトからのリンク）
- (4) オルトメトリクス（SNS やブログの言及数）
- (5) 引用数（データを利用した文献数）
- (6) その他 [具体的に：]
- (7) **あてはまるものはない（関心なし）**

Q33. カレントデータを整備・公開する上で、より詳しく知りたいと思われる項目をお選び下さい [複数選択可]

- (1) 適切なデータ形式
- (2) 適切なリポジトリ
- (3) 適切なメタデータ（作成者やキーワードなど、データを検索するために付与するデータの要約情報）標準（例：ジャパンリンクセンターによる研究データのメタデータ仕様（PDF））
- (4) データのバージョン管理方法
- (5) データのバックアップ方法
- (6) データの安全な管理方法（セキュリティ）

- (7) データに含まれる機密情報の処理
- (8) 知的財産権やライセンス
- (9) その他〔具体的に： 〕
- (10) あてはまるものはない

Q34. カレントデータの整備や公開を、ご自身や共同研究者にかわって図書館員やデータキュレーターなどの第三者が行う場合、分野の知識や専門性が必要であると考えられる項目をお選び下さい。カレントデータが多様な場合は、もっとも難しいと考えられるデータについてお答え下さい。〔複数選択可〕

- (1) 適切なデータ形式への変換
- (2) 適切なリポジトリの選択
- (3) 適切なメタデータ（作成者やキーワードなど、データの要約情報）標準の選択
- (4) メタデータの作成
- (5) データを再利用しやすいように整える
- (6) 機関のリポジトリによるデータ公開
- (7) データを異分野の研究者に紹介する
- (8) その他〔具体的に： 〕
- (9) あてはまるものはない（いずれも専門性を必要としない）

7. 自由回答

Q35. 論文やデータの公開、オープンサイエンス、および調査に関するご意見やご感想がありましたら、ご自由にお書き下さい。（最大 1,000 文字）

ご協力ありがとうございました。

[付録 3] 単純集計結果

1. 研究分野について

Q1. ご自身の研究分野に最も近いものをお選び下さい。＊

付表 2 回答者の研究分野

研究分野	人数	比率
工学	589	41.9%
生物科学	180	12.8%
化学	134	9.5%
農学	123	8.7%
医学	117	8.3%
計算機科学	71	5.0%
物理学	63	4.5%
地球科学	55	3.9%
社会科学	23	1.6%
数学	18	1.3%
心理学	16	1.1%
天文学	4	0.3%
人文学	2	0.1%
その他	3	0.2%
研究を行ったことはない（口頭発表や論文出版の経験はない）	8	0.6%
合計	1,406	100.0%

2. 学術論文について

Q2. ご自身の論文について、あてはまるものをお選び下さい。＊

付表 3 オープンアクセスの状況

OA の論文がある	991	70.9%
OA の論文はない	347	24.8%
わからない	60	4.3%
合計	1,398	100.0%

Q3. 論文をオープンアクセスにした理由として、あてはまるものをお選び下さい。[複数選択可] *

付表 4 論文を OA にした理由 (n=991, 複数回答)

理由	人数	比率
論文を投稿した雑誌がオープンアクセスだから	810	81.7%
研究成果を広く認知してもらいたいから	465	46.9%
科学研究や成果実装を推進したいから	144	14.5%
他の研究者からのリクエストに応じて	101	10.2%
オープンアクセスに貢献したいから	94	9.5%
所属機関のポリシーだから	82	8.3%
分野・コミュニティの規範だから	29	2.9%
助成機関のポリシー（助成条件）だから	20	2.0%
その他	24	2.4%
あてはまるものはない	1	0.1%

Q4. 論文がオープンアクセスではない理由として、あてはまるものをお選び下さい。[複数選択可] *

付表 5 論文が OA ではない理由 (n=407, 複数回答)

理由	人数	比率
投稿したい雑誌がオープンアクセスではないから	246	60.4%
資金が必要だから	161	39.6%
所属機関にオープンアクセス方針がないから	47	11.5%
分野・コミュニティで推奨されていないから	13	3.2%
時間が必要だから	10	2.5%
リポジトリなどの公開手段がないから	9	2.2%
ニーズがないと思うから	9	2.2%
助成機関のポリシー（助成条件）ではないから	4	1.0%
その他	24	5.9%
あてはまるものはない	28	6.9%

Q5. Q4 の理由が解決された場合、論文をオープンアクセスにしたいと思われますか？
＊

付表 6 オープンアクセスの意思

はい	296	78.1%
いいえ	18	4.7%
わからない	65	17.2%
合計	379	100.0%

Q6. 研究に利用したことがある（アイデアの参考にしたり引用した経験がある）論文の分野をお選び下さい。[複数選択可] ＊

付表 7 研究に利用したことがある分野（n=1,398, 複数回答）

研究分野	人数	比率
工学	847	60.6%
生物科学	511	36.6%
化学	473	33.8%
物理学	431	30.8%
医学	381	27.3%
農学	287	20.5%
コンピュータサイエンス	268	19.2%
地球科学	149	10.7%
数学	140	10.0%
社会科学	106	7.6%
心理学	71	5.1%
天文学	33	2.4%
人文学	24	1.7%
その他	5	0.4%

Q7. 論文を探す際に、よく利用する検索ツールや情報源をお選び下さい。[複数選択可]

付表 8 論文の探索ツール (n=1,398, 複数回答)

論文の探索ツール	人数	比率
サーチエンジン (Google, Google Scholar など)	1,164	83.3%
論文情報のデータベース (Web of Science, Scopus, CiNii Articles など)	990	70.8%
論文や学術記事の参考文献	855	61.2%
出版社や学術雑誌のサイト (Elsevier, Wiley など)	683	48.9%
論文やプレプリントのサーバ (PubMed Central, arXiv, J-Stage など)	675	48.3%
研究者や同僚に尋ねる／教えてもらう	469	33.5%
学術機関のリポジトリ・アーカイブ (大学や NASA の機関リポジトリなど)	220	15.7%
学術系 SNS (Mendeley, ResearchGate など)	134	9.6%
アラートサービス (RSS など)	104	7.4%
政府・機関・出版社などの広報, ニュースレター	98	7.0%
ブログや一般的な SNS (Facebook, Twitter など)	72	5.2%
メーリングリスト	64	4.6%
その他	11	0.8%
無回答	4	0.3%

Q8. 論文を利用する際に、その信頼性の判断基準としている項目をお選び下さい。[複数選択可]

付表 9 論文の信頼性の判断基準 (n=1,398, 複数回答)

判断基準	人数	比率
掲載雑誌	1,219	87.2%
著者情報 (所属機関, 職位など)	932	66.7%
研究手法の確かさ	901	64.4%
引用数 (その論文を引用した文献数)	479	34.3%
抄録	227	16.2%
ダウンロード数	42	3.0%
オルトメトリクス (SNS やブログの言及数)	5	0.4%
その他	59	4.2%
無回答	4	0.3%

3. 研究データの提供について

Q9. 共同研究者を除く他の研究者にデータを提供したご経験はありますか？

付表 10 データ提供経験

頻度	人数	比率
よくある	104	7.4%
たまにある	511	36.6%
ほとんどない	469	33.5%
まったくない	303	21.7%
わからない	10	0.7%
無回答	1	0.1%
合計	1,398	100.0%

Q10. 共同研究者を除く他の研究者からデータの提供を受けたご経験はありますか？

付表 11 データ被提供経験

頻度	人数	比率
よくある	75	5.4%
たまにある	466	33.3%
ほとんどない	490	35.1%
まったくない	359	25.7%
わからない	7	0.5%
無回答	1	0.1%
合計	1,398	100.0%

4. 公開データの利用について

Q11. これまでに、公開データを以下の公開先から入手したご経験はありますか？ [複数選択可] *

付表 12 公開データの入手経験

入手経験	人数	比率
あり（チェックあり）	1,060	75.8%
なし	318	22.7%
わからない	20	1.4%
合計	1,398	100.0%

付表 13 公開データの入手方法（n=1,060, 複数回答）

入手方法	人数	比率
個人や研究室のウェブサイト	687	64.8%
論文の補足資料	563	53.1%
学術機関のリポジトリ・データアーカイブ（大学や NASA のリポジトリなど）	530	50.0%
特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）	245	23.1%
学術系 SNS（Mendeley, ResearchGate など）	157	14.8%
コード共有サービス（GitHub など）	113	10.7%
データ共有サービス（figshare, zenodo など）	34	3.2%
その他	13	1.2%

Q12. 公開データの入手にあたって、問題だと感じたことがある項目をお選び下さい。[複数選択可] *

付表 14 公開データ入手における問題の有無

データ入手の問題	人数	比率
あり（チェックあり）	846	79.8%
なし（問題を感じたことはない）	214	20.2%
合計	1,060	100.0%

付表 15 公開データを入手する際の障壁（n=846、複数回答）

公開データの入手障壁	人数	比率
利用料金が必要	365	43.1%
利用者登録が必要	282	33.3%
利用条件がよくわからない	280	33.1%
データごとに品質が異なる	272	32.2%
データごとにフォーマットが異なる	237	28.0%
利用したいデータへのアクセス方法がよくわからない	195	23.0%
著作者情報がよくわからない	172	20.3%
データの解釈や再利用方法がよくわからない	140	16.5%
リクエストから入手までに時間がかかる	139	16.4%
最新のデータを入手できない	120	14.2%
その他	18	2.1%

Q13. 入手した公開データの利用目的について、あてはまるものをそれぞれお選び下さい。
*

付表 16 公開データの利用目的（n=1,060）

	行う		たまに行う		行わない	
	人数	比率	人数	比率	人数	比率
研究のアイデアや仮説の参考にする	378	35.7%	589	55.6%	93	8.8%
再分析・再利用して自身の研究を行う	176	16.6%	410	38.7%	474	44.7%
研究を再現・追試する	90	8.5%	399	37.6%	571	53.9%

Q14. 研究に利用したことがある（参考にしたり再分析した経験がある）公開データの分野をお選び下さい。[複数選択可] ＊

付表 17 公開データを研究に利用した経験の有無

利用経験	人数	比率
あり（チェックあり）	1,034	97.5%
なし（公開データを入手しても、研究に利用したことはない）	26	2.5%
合計	1,060	100.0%

付表 18 研究に利用したことがある公開データの分野（n=1,034, 複数回答）

研究分野	人数	比率
工学	518	50.1%
生物科学	348	33.7%
化学	271	26.2%
物理学	208	20.1%
医学	202	19.5%
コンピュータサイエンス	163	15.8%
農学	160	15.5%
地球科学	91	8.8%
数学	48	4.6%
社会科学	39	3.8%
心理学	21	2.0%
天文学	19	1.8%
人文学	14	1.4%
その他	2	0.2%

Q15. 公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい。[複数選択可]

付表 19 公開データの探索ツール (n=1,034, 複数回答)

データの探索ツール	人数	比率
サーチエンジン (Google など)	864	83.6%
論文や学術記事の参考文献	734	71.0%
出版社や学術雑誌のサイト (Elsevier, Wiley など)	391	37.8%
研究者や同僚に尋ねる／教えてもらう	375	36.3%
学術機関のリポジトリ・データアーカイブ	338	32.7%
データ情報のデータベース (Data Citation Index など)	187	18.1%
学術系 SNS (Mendeley, ResearchGate など)	97	9.4%
データジャーナル (簡易なデータ記述とデータへのリンクを掲載した雑誌)	79	7.6%
政府・機関・出版社などの広報, ニュースレター	77	7.4%
メーリングリスト	36	3.5%
ブログや一般的な SNS	30	2.9%
特定のデータリポジトリ	27	2.6%
アラートサービス (RSS など)	19	1.8%
その他	5	0.5%
無回答	1	0.1%

Q16. 公開データを利用する際に、信頼性の判断基準としている項目をお選び下さい。[複数選択可]

付表 20 公開データの信頼性の判断基準 (n=1,034, 複数回答)

判断基準	人数	比率
著者情報 (所属機関, 職位など)	733	70.9%
研究手法の確かさ	649	62.8%
そのデータを用いた論文	602	58.2%
掲載リポジトリ	264	25.5%
引用数 (そのデータを引用した文献数)	237	22.9%
データにつけられた説明 (Readme ファイルなど)	221	21.4%
ダウンロード数	54	5.2%
オルトメトリクス (SNS やブログの言及数)	9	0.9%
その他	20	1.9%
無回答	9	0.9%

Q17. 今後、利用してみたいと思う公開データの分野をお選び下さい。[複数選択可] *

付表 21 利用したい公開データの有無

利用したい公開データ	人数	比率
あり（チェックあり）	1,357	97.1%
他の研究者の公開データを利用したいとは思わない	41	2.9%
合計	1,398	100.0%

付表 22 今後、利用してみたいと思う公開データの分野（n=1,357，複数回答）

研究分野	人数	比率
工学	767	56.5%
生物科学	515	38.0%
化学	434	32.0%
医学	400	29.5%
物理学	394	29.0%
コンピュータサイエンス	342	25.2%
農学	303	22.3%
地球科学	171	12.6%
社会科学	143	10.5%
数学	136	10.0%
心理学	89	6.6%
天文学	42	3.1%
人文学	39	2.9%
その他	8	0.6%

5. 研究データの公開について

Q18. これまでに、研究データを以下の方法で公開したご経験はありますか？〔複数選択可〕＊

付表 23 データ公開経験の有無

データ公開経験	人数	比率
あり（チェックあり）	713	51.0%
なし	656	46.9%
わからない	27	1.9%
データは用いない	2	0.1%
合計	1,398	100.0%

付表 24 データを公開した方法（n=713, 複数回答）

公開方法	人数	比率
個人や研究室のウェブサイト	362	50.8%
論文の補足資料（supplementary materials）	335	47.0%
所属機関のリポジトリ・データアーカイブ	244	34.2%
特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）	117	16.4%
学術系 SNS（Mendeley, ResearchGate など）	70	9.8%
コード共有サービス（GitHub など）	24	3.4%
データ共有サービス（figshare, zenodo など）	16	2.2%
その他	18	2.5%

Q19. 研究データを公開した理由として、あてはまるものをお選び下さい。[複数選択可]

＊

付表 25 データを公開した理由 (n=713, 複数回答)

理由	人数	比率
研究成果を広く認知してもらいたいから	417	58.5%
論文を投稿した雑誌のポリシー（投稿規定）だから	312	43.8%
科学研究や成果実装を推進したいから	186	26.1%
他の研究者からのリクエストに応じて	181	25.4%
所属機関のポリシーだから	169	23.7%
オープンデータに貢献したいから	75	10.5%
分野・コミュニティの規範だから	68	9.5%
業績になる場合があるから	66	9.3%
助成機関のポリシー（助成条件）だから	45	6.3%
その他	14	2.0%
あてはまるものはない	5	0.7%

Q20. 研究データを公開していない理由として、あてはまるものをお選び下さい。[複数選択可] ＊

付表 26 データを公開していない理由 (n=683, 複数回答)

理由	人数	比率
論文を投稿した雑誌のポリシー（投稿条件）ではないから	180	26.4%
時間が必要だから	177	25.9%
所属機関にデータ公開ポリシーがないから	153	22.4%
ニーズがないと思うから	151	22.1%
業績にならないから	148	21.7%
リポジトリなどの公開手段がないから	61	8.9%
資金が必要だから	52	7.6%
助成機関のポリシー（助成条件）ではないから	42	6.1%
分野・コミュニティで推奨されていないから	29	4.2%
その他	61	8.9%
あてはまるものはない	88	12.9%

Q21.Q20 の理由が解決された場合、研究データを公開したいと思いますか？＊

付表 27 データ公開意思

データ公開意思	人数	比率
はい	169	28.4%
いいえ	130	21.8%
わからない	296	49.7%
合計	595	100.0%

Q22.所属機関では、研究データの保存期間は定められていますか？

付表 28 データ保存期間規定の有無

保存規定	人数	比率
定められている	672	48.1%
定められていない	380	27.2%
わからない	344	24.6%
合計	1,396	100.0%

6. 最近の研究のためのデータについて

Q23.カレントデータの総量は、およそどれくらいでしたか？ 論文などには使用しなかったデータも含めてあてはまる単位をお選びください。

付表 29 カレントデータの量

データの量	人数	比率
MB（メガバイト）以下	262	18.8%
GB（ギガバイト）	781	55.9%
TB（テラバイト）	188	13.5%
PB（ペタバイト）以上	4	0.3%
わからない	158	11.3%
無回答	3	0.2%
合計	1,396	100.0%

Q24. ご自身以外で、カレントデータの所有権をもつ人・組織をお選び下さい。[複数選択可]

付表 30 カレントデータの所有権をもつ人・組織の有無

所有権をもつ人・組織	人数	比率
あり（チェックあり）	1,166	83.5%
なし（自分のみ）	115	8.2%
わからない	114	8.2%
無回答	1	0.1%
合計	1,396	100.0%

付表 31 カレントデータの所有権をもつ人・組織（n=1,166, 複数回答）

所有権をもつ人・組織	人数	比率
共同研究者	957	82.1%
所属機関	679	58.2%
所属機関以外の研究機関	111	9.5%
所属機関以外の企業	75	6.4%
研究助成機関	67	5.7%
その他	7	0.6%

Q25. カレントデータには、以下の情報が含まれていますか？ [複数選択可]

付表 32 カレントデータの機密情報の有無

機密情報	人数	比率
あり（チェックあり）	683	48.9%
なし	493	35.3%
わからない	207	14.8%
無回答	13	0.9%
合計	1,396	100.0%

付表 33 カレントデータに含まれる機密情報（n=683, 複数回答）

機密情報	人数	比率
個人情報	293	42.9%
企業・商業上の機密情報	288	42.2%
健康情報（遺伝情報, 医療情報）	76	11.1%
その他, 守秘義務がある情報	399	58.4%

Q26. カレントデータは何年くらい保存する必要があると思われますか？ 年数を数字で入力して下さい。(保存の必要はないと考える場合は「0」、永久保存の場合は「999」と入力して下さい)

付表 34 カレントデータの望ましい保存期間

望ましい保存期間	人数	比率
0 年 (保存の必要はない)	42	3.0%
1 年	6	0.4%
2 年	14	1.0%
3 年	78	5.6%
4 年	2	0.1%
5 年	422	30.2%
6 年	0	0.0%
7 年	11	0.8%
8 年	4	0.3%
9 年	1	0.1%
10 年	432	30.9%
12 年	1	0.1%
15 年	23	1.6%
20 年	81	5.8%
25 年	5	0.4%
30 年	29	2.1%
40 年	1	0.1%
50 年	23	1.6%
99 年	2	0.1%
100 年	5	0.4%
永久保存	190	13.6%
無回答	24	1.7%
合計	1,396	100.0%

Q27.現在までに、カレントデータ（データの一部）を公開しましたか？＊

付表 35 カレントデータ公開の有無

カレントデータの公開	人数	比率
はい	254	18.2%
いいえ	1,005	72.0%
わからない	137	9.8%
合計	1,396	100.0%

Q28.カレントデータを管理・公開しようとする場合、次の資源は十分に整っていますか（いましたか）？＊

付表 36 カレントデータの管理資源の充実度

	0		1		2		3		4		合計	
	人	%	人	%	人	%	人	%	人	%	人	%
研究中のデータ 用ストレージ	197	14.1	429	30.7	238	17.0	387	27.7	145	10.4	1,396	100.0
データ公開用の リポジトリ	378	27.1	534	38.3	228	16.3	195	14.0	61	4.4	1,396	100.0
研究終了後のデ ータ保存用スト レージ	201	14.4	521	37.3	277	19.8	289	20.7	108	7.7	1,396	100.0
データの整備・公 開のための時間	239	17.1	782	56.0	258	18.5	96	6.9	21	1.5	1,396	100.0
データの整備・公 開のための人材	216	15.5	922	66.0	183	13.1	61	4.4	14	1.0	1,396	100.0
データの整備・公 開のための資金	262	18.8	834	59.7	201	14.4	77	5.5	22	1.6	1,396	100.0

0=わからない, 1=不十分, 2=やや不十分, 3=ほぼ充分, 4=充分

Q29. カレントデータを論文の発表前に公開しようとする場合、次の点は問題（懸案）と
なりますか（なりましたか）？＊

付表 37 カレントデータを公開する際の障壁

	0		1		2		3		4		合計	
	人	%	人	%	人	%	人	%	人	%	人	%
商用利用される可能性	106	7.6	148	10.6	166	11.9	330	23.6	646	46.3	1,396	100.0
機密・プライバシー情報	85	6.1	156	11.2	151	10.8	309	22.1	695	49.8	1,396	100.0
誤解や誤用の可能性	82	5.9	100	7.2	181	13.0	449	32.2	584	41.8	1,396	100.0
引用せずに利用される可能性	67	4.8	26	1.9	78	5.6	321	23.0	904	64.8	1,396	100.0
研究の誤りを発見される可能性	90	6.4	526	37.7	576	41.3	122	8.7	82	5.9	1,396	100.0
公開したデータを使って自分より先に論文を出版される可能性	66	4.7	50	3.6	99	7.1	295	21.1	886	63.5	1,396	100.0

0=わからない, 1=問題ではない, 2=あまり問題ではない, 3=やや問題である, 4=問題である

Q30. カレントデータを公開した場合（公開したカレントデータは）、ご自身と同じ分野、
すなわち Q1 で選択した分野の研究者の多くが理解できると思われますか？ ＊

付表 38 カレントデータの理解：同じ分野

カレントデータの理解	人数	比率
できると思う	838	60.0%
やや難しいと思う	325	23.3%
難しいと思う	145	10.4%
わからない	88	6.3%
合計	1,396	100.0%

Q31. カレントデータを公開した場合（公開したカレントデータは）、異分野の研究者の多くが理解できると思われますか？＊

付表 39 カレントデータの理解：他分野

カレントデータの理解	人数	比率
できると思う	184	13.2%
やや難しいと思う	522	37.4%
難しいと思う	597	42.8%
わからない	93	6.7%
合計	1,396	100.0%

Q32. カレントデータを公開した場合（公開したカレントデータについて）、ご関心がある項目をお選び下さい。[複数選択可]

付表 40 カレントデータのメトリクスへの関心

メトリクスへの関心	人数	比率
チェックあり（関心あり）	1,198	85.8%
あてはまるものはない（関心なし）	195	14.0%
無回答	3	0.2%
合計	1,396	100.0%

付表 41 関心のあるメトリクス（n=1,198, 複数回答）

メトリクス	人数	比率
引用数（データを利用した文献数）	874	73.0%
閲覧数	858	71.6%
ダウンロード数	805	67.2%
リンク数（別のウェブサイトからのリンク）	270	22.5%
オルトメトリクス（SNS やブログの言及数）	122	10.2%
その他	32	2.7%

Q33. カレントデータを整備・公開する上で、より詳しく知りたいと思われる項目をお選び下さい [複数選択可]

付表 42 カレントデータの管理方法への関心

管理方法への関心	人数	比率
チェックあり（詳しく知りたい項目あり）	1,257	90.0%
あてはまるものはない	137	9.8%
無回答	2	0.1%
合計	1,396	100.0%

付表 43 データの整備・公開に関して詳しく知りたい項目（n=1,257、複数回答）

項目	人数	比率
知的財産権やライセンス	869	69.1%
データの安全な管理方法（セキュリティ）	836	66.5%
適切なデータ形式	754	60.0%
データに含まれる機密情報の処理	697	55.4%
適切ななりポジトリ	610	48.5%
適切なメタデータ*標準	494	39.3%
データのバックアップ方法	410	32.6%
データのバージョン管理方法	407	32.4%
その他	16	1.3%

*作成者やキーワードなど、データを検索するために付与するデータの要約情報（例：ジャパンリンクセンターによる研究データのメタデータ仕様（PDF））

Q34. カレントデータの整備や公開を、ご自身や共同研究者にかわって図書館員やデータキュレーターなどの第三者が行う場合、分野の知識や専門性が必要であると考えられる項目をお選び下さい。カレントデータが多様な場合は、もっとも難しいと考えられるデータについてお答え下さい。[複数選択可]

付表 44 第三者によるカレントデータの管理における専門性

カレントデータの整備や公開における専門性	人数	比率
チェックあり（専門性を必要とする）	1,302	93.3%
あてはまるものはない（いずれも専門性を必要としない）	93	6.7%
無回答	1	0.1%
合計	1,396	100.0%

付表 45 カレントデータの管理に専門性を必要とする項目（n=1,302, 複数回答）

項目	人数	比率
適切なデータ形式への変換	781	60.0%
データを再利用しやすいように整える	769	59.1%
適切なメタデータ標準の選択	656	50.4%
適切なリポジトリの選択	531	40.8%
メタデータの作成	445	34.2%
データを異分野の研究者に紹介する	426	32.7%
機関のリポジトリによるデータ公開	347	26.7%
その他	31	2.4%

7. 自由回答

Q35. 論文やデータの公開、オープンサイエンス、および調査に関するご意見や感想がありましたら、ご自由にお書き下さい。（最大 1,000 文字）

※自由回答は報告書の電子付録として公開している（池内ら, 2017）。

[付録 4] 回答者の分野と年齢

付表 46 回答者の分野と年齢

分野	21-30 歳		31-40 歳		41-50 歳		51-60 歳		61-70 歳		合計	
	人	%	人	%	人	%	人	%	人	%	人	%
農業科学	0	0.0	27	58.7	7	15.2	9	19.6	3	6.5	46	100.0
生物学	2	2.4	40	47.1	29	34.1	12	14.1	2	2.4	85	100.0
化学	4	3.3	69	57.0	35	28.9	11	9.1	2	1.7	121	100.0
臨床医学	0	0.0	3	13.6	11	50.0	2	9.1	6	27.3	22	100.0
CS	1	1.6	29	46.0	15	23.8	14	22.2	4	6.3	63	100.0
経済学	0	0.0	3	42.9	0	0.0	3	42.9	1	14.3	7	100.0
工学	9	1.9	203	43.5	125	26.8	70	15.0	60	12.8	467	100.0
環境学	2	4.3	19	41.3	14	30.4	7	15.2	4	8.7	46	100.0
地球科学	1	2.4	18	42.9	16	38.1	6	14.3	1	2.4	42	100.0
免疫学	0	0.0	2	16.7	5	41.7	2	16.7	3	25.0	12	100.0
材料科学	3	2.5	48	40.7	41	34.7	17	14.4	9	7.6	118	100.0
数学	0	0.0	6	40.0	3	20.0	4	26.7	2	13.3	15	100.0
微生物学	0	0.0	23	47.9	16	33.3	6	12.5	3	6.3	48	100.0
分子生物学	1	2.3	21	47.7	14	31.8	3	6.8	5	11.4	44	100.0
複合領域	2	3.0	31	47.0	24	36.4	5	7.6	4	6.1	66	100.0
神経科学	3	6.5	26	56.5	11	23.9	3	6.5	3	6.5	46	100.0
薬理学	0	0.0	22	62.9	7	20.0	2	5.7	4	11.4	35	100.0
物理学	2	3.9	26	51.0	15	29.4	6	11.8	2	3.9	51	100.0
植物動物学	0	0.0	10	52.6	8	42.1	1	5.3	0	0.0	19	100.0
精神医学	0	0.0	9	69.2	4	30.8	0	0.0	0	0.0	13	100.0
社会科学	0	0.0	2	20.0	4	40.0	2	20.0	2	20.0	10	100.0
宇宙科学	0	0.0	4	50.0	3	37.5	0	0.0	1	12.5	8	100.0
その他	0	0.0	6	42.9	4	28.6	3	21.4	1	7.1	14	100.0
合計	30	2.1	647	46.3	411	29.4	188	13.4	122	8.7	1,398	100.0

[付録 5] 分野別同じ分野の研究者によるデータの理解の難しさ

付表 47 分野別同じ分野の研究者によるデータの理解の難しさ (n=1,382)

分野	n	難しいと思う	やや難しい と思う	できると思う	わからない
宇宙科学	8	0.0%	12.5%	75.0%	12.5%
社会科学	10	10.0%	10.0%	60.0%	20.0%
精神医学	13	15.4%	30.8%	53.8%	0.0%
植物動物学	19	5.3%	36.8%	57.9%	0.0%
物理学	51	11.8%	27.5%	58.8%	2.0%
薬理学	35	14.3%	5.7%	74.3%	5.7%
神経科学	46	15.2%	28.3%	54.3%	2.2%
複合領域	66	10.6%	24.2%	59.1%	6.1%
分子生物学	44	11.4%	20.5%	65.9%	2.3%
微生物学	47	12.8%	10.6%	68.1%	8.5%
数学	15	33.3%	20.0%	33.3%	13.3%
材料科学	118	15.3%	21.2%	60.2%	3.4%
免疫学	12	0.0%	41.7%	50.0%	8.3%
地球科学	42	0.0%	31.0%	66.7%	2.4%
環境学	46	4.3%	23.9%	65.2%	6.5%
工学	467	11.6%	24.6%	57.6%	6.2%
経済学	7	0.0%	0.0%	85.7%	14.3%
CS	63	6.3%	23.8%	57.1%	12.7%
臨床医学	22	4.5%	36.4%	50.0%	9.1%
化学	120	10.0%	20.0%	62.5%	7.5%
生物学	85	4.7%	28.2%	61.2%	5.9%
農業科学	46	6.5%	15.2%	69.6%	8.7%

[付録 6] 所属機関別データ公開の障壁の程度

付表 48 所属機関別「誤解や誤用の可能性」への懸念

所属	問題である	やや問題で ある	あまり問題 ではない	問題では ない	わからない	回答者数
大学	371 44.0%	265 31.4%	109 12.9%	58 6.9%	40 4.7%	843 100.0%
公的機関・ 団体	85 38.1%	71 31.8%	27 12.1%	24 10.8%	16 7.2%	223 100.0%
企業	124 38.5%	110 34.2%	44 13.7%	18 5.6%	26 8.1%	322 100.0%
合計	580 41.8%	446 32.1%	180 13.0%	100 7.2%	82 5.9%	1,388 100.0%

[付録 7] 所属機関別カレントデータの理解の難しさ

付表 49 所属機関別カレントデータの理解（自分野）

所属	できると思う	やや難しいと 思う	難しいと思う	わからない	回答者数
大学	504 59.8%	195 23.1%	97 11.5%	47 5.6%	843 100.0%
公的機関・団体	138 61.9%	55 24.7%	19 8.5%	11 4.9%	223 100.0%
企業	192 59.6%	72 22.4%	29 9.0%	29 9.0%	322 100.0%
合計	834 60.1%	322 23.2%	145 10.4%	87 6.3%	1,388 100.0%

付表 50 所属機関別カレントデータの理解（他分野）

所属	できると思う	やや難しいと思う	難しいと思う	わからない	回答者数
大学	108 12.8%	313 37.1%	372 44.1%	50 5.9%	843 100.0%
公的機関・団体	33 14.8%	89 39.9%	89 39.9%	12 5.4%	223 100.0%
企業	42 13.0%	115 35.7%	135 41.9%	30 9.3%	322 100.0%
合計	183 13.2%	517 37.2%	596 42.9%	92 6.6%	1,388 100.0%

[付録 8] 国別データリポジトリ数

付表 51 国別データリポジトリ数

分野	リポジトリ数	比率
United States	1,012	33.0%
Germany	337	11.0%
United Kingdom	294	9.6%
International	202	6.6%
European Union	199	6.5%
Canada	184	6.0%
France	98	3.2%
Australia	85	2.8%
Switzerland	68	2.2%
Japan	57	1.9%
Netherlands	52	1.7%
India	48	1.6%
China	38	1.2%
Austria	34	1.1%
Italy	32	1.0%
その他（58 か国）	323	10.5%
合計	3,063	100.0%

出典：re3data.org を用いて著者作成（2018 年 9 月 19 日現在）

[付録 9] IF の分野内順位とポリシーの強度

付表 52 IF の分野内順位とポリシーの強度 (n=220)

順位	リポジトリ (RP)				補足資料 (SP)			
	必須	推奨	受諾	なし	必須	推奨	受諾	なし
1	14	2	0	6	8	5	8	1
2	15	2	0	5	7	1	11	3
3	11	2	0	9	5	2	12	3
4	9	3	0	10	2	3	15	2
5	12	2	0	8	2	6	12	2
6	9	2	0	11	2	1	17	2
7	6	6	0	10	2	6	12	2
8	8	5	0	9	0	1	20	1
9	6	3	1	12	3	1	13	5
10	10	3	0	9	0	3	17	2
合計	100	30	1	89	31	29	137	23

[付録 10] 刊行開始年代とポリシーの強度

付表 53 刊行開始年代とポリシーの強度

年代	雑誌数	累積	リポジトリ (RP)				補足資料 (SP)			
			必須	推奨	受諾	なし	必須	推奨	受諾	なし
2010	8	3.6%	6	1	0	1	3	2	3	0
			75.0%	12.5%	0.0%	12.5%	37.5%	25.0%	37.5%	0.0%
2000	54	28.2%	34	4	0	16	12	11	28	3
			63.0%	7.4%	0.0%	29.6%	22.2%	20.4%	51.9%	5.6%
1990	36	44.5%	17	6	1	12	6	3	25	2
			47.2%	16.7%	2.8%	33.3%	16.7%	8.3%	69.4%	5.6%
1980	33	59.5%	14	6	0	13	4	1	24	4
			42.4%	18.2%	0.0%	39.4%	12.1%	3.0%	72.7%	12.1%
1970	18	67.7%	3	3	0	12	0	2	14	2
			16.7%	16.7%	0.0%	66.7%	0.0%	11.1%	77.8%	11.1%
1960	12	73.2%	2	2	0	8	0	2	9	1
			16.7%	16.7%	0.0%	66.7%	0.0%	16.7%	75.0%	8.3%
1950	13	79.1%	3	3	0	7	2	3	5	3
			23.1%	23.1%	0.0%	53.8%	15.4%	23.1%	38.5%	23.1%
-1940	46	100.0%	21	5	0	20	4	5	29	8
			45.7%	10.9%	0.0%	43.5%	8.7%	10.9%	63.0%	17.4%
合計	220		100	30	1	89	31	29	137	23

[付録 11] 雑誌年齢の分野内順位とポリシーの強度

付表 54 雑誌年齢の分野内順位とポリシーの強度 (n=220)

順位	リポジトリ (RP)				補足資料 (SP)			
	必須	推奨	受諾	なし	必須	推奨	受諾	なし
1	11	3	0	8	5	5	10	2
2	13	1	0	8	5	3	13	1
3	12	2	0	8	4	2	15	1
4	10	5	0	7	2	2	14	4
5	7	5	0	10	2	3	15	2
6	7	2	0	13	3	3	14	2
7	7	6	1	8	1	5	12	4
8	12	2	0	8	3	2	14	3
9	12	2	0	8	6	2	12	2
10	9	2	0	11	0	2	18	2
合計	100	30	1	89	31	29	137	23